BỘ CÔNG THƯƠNG

**TRƯỜNG ĐẠI HỌC CÔNG THƯƠNG TP.HCM**

**KHOA: CÔNG NGHỆ THÔNG TIN**

A blue and black logo

AI-generated content may be incorrect.

**ĐỒ ÁN CUỐI KỲ**

**DỮ LIỆU NOSQL**

**ĐỀ TÀI: XÂY DỰNG WEBSITE QUẢN LÝ THÔNG TIN TUYỂN DỤNG**

|  |  |  |
| --- | --- | --- |
| **GDHD** | **Nguyễn Thị Thu Tâm** | |
| **NHÓM** | **11** | |
| **THÀNH VIÊN** | **Đặng Trường Vũ** | **2001225908** |
|  | **Đoàn Duy Hiếu** | **2001221914** |
|  | **Lê Trương Công Hiếu** | **2001221414** |

**THÀNH PHỐ HỒ CHÍ MINH, THÁNG … NĂM …**

bảng phân công công việc

|  |  |  |  |
| --- | --- | --- | --- |
| MSSV | Họ và tên | Công việc | Tiến độ hoàn thành |
| 2001225908 | Đặng Trường Vũ |  |  |
| 2001221914 | Đoàn Duy Hiếu |  |  |
| 2001221414 | Lê Trương Công Hiếu |  |  |

Mục lục

[CHƯƠNG 1: Tổng quan về Redis 1](#_Toc211805312)

[1.1 Giới thiệu Redis 1](#_Toc211805313)

[1.1.1 Redis là gì? 1](#_Toc211805314)

[1.1.2 Các tính năng nổi bật của Redis 1](#_Toc211805315)

[1.1.3 Khi nào nên sử dụng Redis? 1](#_Toc211805316)

[1.1.4 Redis so với Cơ sở dữ liệu truyền thống 2](#_Toc211805317)

[1.2 Kiến trúc 3](#_Toc211805318)

[1.2.1 Thành phần Client-Server 3](#_Toc211805319)

[1.2.2 Persistence (Tính bền vững) 4](#_Toc211805320)

[1.2.3 Replication (Nhân bản) 6](#_Toc211805321)

[1.2.4 Cluster 8](#_Toc211805322)

[1.2.5 Sentinel 9](#_Toc211805323)

[1.2.6 Memory Management (Quản lý bộ nhớ) 11](#_Toc211805324)

[1.3 Hướng dẫn cài đặt 12](#_Toc211805325)

[1.3.1 Cài đặt Redis trên Windows 12](#_Toc211805326)

[1.3.2 Cài đặt Redis thông qua WSL (Windows Subsystem for Linux) 16](#_Toc211805327)

[1.3.3 Các công cụ GUI tiêu biểu cho Redis 18](#_Toc211805328)

[1.4 Cấu trúc dữ liệu trong Redis 19](#_Toc211805329)

[1.4.1 Strings 19](#_Toc211805330)

[1.4.1.1 Giới thiệu 19](#_Toc211805331)

[1.4.1.2 Lệnh cơ bản 19](#_Toc211805332)

[1.4.1.3 Ví dụ minh họa 20](#_Toc211805333)

[1.4.1.4 Ứng dụng thực tế 21](#_Toc211805334)

[1.4.2 Lists 21](#_Toc211805335)

[1.4.2.1 Giới thiệu 21](#_Toc211805336)

[1.4.2.2 Lệnh cơ bản 22](#_Toc211805337)

[1.4.2.3 Ví dụ minh họa 22](#_Toc211805338)

[1.4.2.4 Ứng dụng thực tế 23](#_Toc211805339)

[1.4.3 Sets 23](#_Toc211805340)

[1.4.3.1 Giới thiệu 23](#_Toc211805341)

[1.4.3.2 Lệnh cơ bản 23](#_Toc211805342)

[1.4.3.3 Ví dụ minh họa 24](#_Toc211805343)

[1.4.3.4 Ứng dụng thực tế 24](#_Toc211805344)

[1.4.4 Hashes 25](#_Toc211805345)

[1.4.4.1 Giới thiệu 25](#_Toc211805346)

[1.4.4.2 Lệnh cơ bản 25](#_Toc211805347)

[1.4.4.3 Ví dụ minh họa 25](#_Toc211805348)

[1.4.4.4 Ứng dụng thực tế 26](#_Toc211805349)

[1.4.5 Sorted Sets 26](#_Toc211805350)

[1.4.5.1 Giới thiệu 26](#_Toc211805351)

[1.4.5.2 Lệnh cơ bản 26](#_Toc211805352)

[1.4.5.3 Ví dụ minh họa 27](#_Toc211805353)

[1.4.5.4 Ứng dụng thực tế 27](#_Toc211805354)

[1.4.6 Bitmap 28](#_Toc211805355)

[1.4.6.1 Giới thiệu 28](#_Toc211805356)

[1.4.6.2 Lệnh cơ bản 28](#_Toc211805357)

[1.4.6.3 Ví dụ minh họa 28](#_Toc211805358)

[1.4.6.4 Ứng dụng thực tế 29](#_Toc211805359)

[1.4.7 HyperLogLog 29](#_Toc211805360)

[1.4.7.1 Giới thiệu 29](#_Toc211805361)

[1.4.7.2 Lệnh cơ bản 30](#_Toc211805362)

[1.4.7.3 Ví dụ minh họa 30](#_Toc211805363)

[1.4.7.4 Ứng dụng thực tế 31](#_Toc211805364)

[1.4.8 Streams 31](#_Toc211805365)

[1.4.8.1 Giới thiệu 31](#_Toc211805366)

[1.4.8.2 Lệnh cơ bản 31](#_Toc211805367)

[1.4.8.3 Ví dụ minh họa 32](#_Toc211805368)

[1.4.8.4 Ứng dụng thực tế 32](#_Toc211805369)

[1.4.9 Geo 33](#_Toc211805370)

[1.4.9.1 Giới thiệu 33](#_Toc211805371)

[1.4.9.2 Lệnh cơ bản 33](#_Toc211805372)

[1.4.9.3 Ví dụ minh họa 33](#_Toc211805373)

[1.4.9.4 Ứng dụng thực tế 34](#_Toc211805374)

[1.4.10 JSON 34](#_Toc211805375)

[1.4.10.1 Giới thiệu 34](#_Toc211805376)

[1.4.10.2 Lệnh cơ bản 34](#_Toc211805377)

[1.4.10.3 Ví dụ minh họa 35](#_Toc211805378)

[1.4.10.4 Ứng dụng thực tế 35](#_Toc211805379)

[1.4.11 TimeSeries 36](#_Toc211805380)

[1.4.11.1 Giới thiệu 36](#_Toc211805381)

[1.4.11.2 Lệnh cơ bản 36](#_Toc211805382)

[1.4.11.3 Ví dụ minh họa 36](#_Toc211805383)

[1.4.11.4 Ứng dụng thực tế 37](#_Toc211805384)

[1.5 Truy vấn Full-Text Search (FT) trong Redis 37](#_Toc211805385)

[1.5.1 Giới thiệu 37](#_Toc211805386)

[1.5.2 Lệnh cơ bản 38](#_Toc211805387)

[1.5.3 Ứng dụng thực tế 39](#_Toc211805388)

[1.5.4 Cách sử dụng FT 39](#_Toc211805389)

[1.5.5 Ví dụ minh họa 41](#_Toc211805390)

[1.6 Một số thao tác quản trị cơ bản 42](#_Toc211805391)

[1.6.1 Kiểm tra thông tin server 42](#_Toc211805392)

[1.6.1.1 Giới thiệu 42](#_Toc211805393)

[1.6.1.2 Lệnh cơ bản 42](#_Toc211805394)

[1.6.1.3 Ứng dụng thực tế 42](#_Toc211805395)

[1.6.2 Quản lý cơ sở dữ liệu 43](#_Toc211805396)

[1.6.2.1 Giới thiệu 43](#_Toc211805397)

[1.6.2.2 Lệnh cơ bản 43](#_Toc211805398)

[1.6.2.3 Ứng dụng thực tế 43](#_Toc211805399)

[1.6.3 Xóa dữ liệu 43](#_Toc211805400)

[1.6.3.1 Giới thiệu 43](#_Toc211805401)

[1.6.3.2 Lệnh cơ bản 44](#_Toc211805402)

[1.6.3.3 Ứng dụng thực tế 44](#_Toc211805403)

[1.6.4 Kiểm tra key và bộ nhớ 44](#_Toc211805404)

[1.7 Kết luận 45](#_Toc211805405)

[CHƯƠNG 2: PHÂN TÍCH VÀ THIẾT KẾ CƠ SỞ DỮ LIỆU NoSQL 46](#_Toc211805406)

[2.1 Bối cảnh 46](#_Toc211805407)

[2.2 Phân tích thị trường hiện tại 46](#_Toc211805408)

[2.3 Giải pháp đề xuất 46](#_Toc211805409)

[2.4 Phân tích yêu cầu hệ thống 47](#_Toc211805410)

[2.4.1 Yêu cầu chức năng 47](#_Toc211805411)

[2.4.2 Yêu cầu phi chức năng 47](#_Toc211805412)

[2.5 Lựa chọn công nghệ: So sánh SQL và NoSQL 48](#_Toc211805413)

[2.5.1 Phân tích đặc thù của hệ thống tuyển dụng 48](#_Toc211805414)

[2.5.2 So sánh SQL và NoSQL trong bối cảnh tuyển dụng 49](#_Toc211805415)

[2.5.3 Hạn chế của NoSQL và cách khắc phục 50](#_Toc211805416)

[2.6 Thiết kế cơ sở dữ liệu NoSQL 51](#_Toc211805417)

[2.6.1 Triết lý thiết kế: Embedded vs Reference 51](#_Toc211805418)

[2.6.2 Mô hình dữ liệu chi tiết 51](#_Toc211805419)

[2.7 So sánh hiệu năng truy vấn 54](#_Toc211805420)

[2.7.1 Kịch bản 1: Hiển thị chi tiết tin tuyển dụng với danh sách ứng viên 54](#_Toc211805421)

[2.7.2 Kịch bản 2: Tìm kiếm công việc theo nhiều tiêu chí 55](#_Toc211805422)

[2.7.3 Kịch bản 3: Cập nhật trạng thái ứng viên 57](#_Toc211805423)

[2.7.4 Tổng kết so sánh hiệu năng 59](#_Toc211805424)

[2.8 Chiến lược caching với Redis 60](#_Toc211805425)

[2.9 Kết luận 60](#_Toc211805426)

[CHƯƠNG 3: TRIỂN KHAI ỨNG DỤNG 62](#_Toc211805427)

[3.1 Tổng quan kiến trúc hệ thống 62](#_Toc211805428)

[3.1.1 Kiến trúc tổng thể 62](#_Toc211805429)

[3.1.2 Luồng hoạt động của hệ thống 62](#_Toc211805430)

[3.2 Công nghệ và công cụ sử dụng 63](#_Toc211805431)

[3.2.1 Backend - ASP.NET Core Web API 63](#_Toc211805432)

[3.2.2 Cơ sở dữ liệu - MongoDB 64](#_Toc211805433)

[3.2.3 Caching và Session - Redis 64](#_Toc211805434)

[3.2.4 Xác thực và bảo mật - JWT 65](#_Toc211805435)

[3.2.5 Các công nghệ hỗ trợ 65](#_Toc211805436)

[3.3 Cài đặt và cấu hình môi trường 66](#_Toc211805437)

[3.3.1 Khởi tạo project ASP.NET Core 66](#_Toc211805438)

[3.3.2 Cấu hình Dependency Injection 66](#_Toc211805439)

[3.3.3 Tích hợp MongoDB Driver 67](#_Toc211805440)

[3.3.4 Tích hợp Redis với StackExchange.Redis 67](#_Toc211805441)

[3.3.5 Cấu hình JWT Authentication 68](#_Toc211805442)

[3.4 Triển khai các API Endpoints 68](#_Toc211805443)

[3.4.1 Authentication APIs 68](#_Toc211805444)

[3.4.2 User Management APIs 69](#_Toc211805445)

[3.4.3 Job Management APIs 70](#_Toc211805446)

[3.4.4 Application APIs 71](#_Toc211805447)

[3.4.5 Company Management APIs 71](#_Toc211805448)

[3.5 Kết luận 73](#_Toc211805449)

[CHƯƠNG 4: Kết luận 75](#_Toc211805450)

MỤC LỤC ẢNH

[Bảng 1.1: Bảng so sánh Redis với RDBMS 3](#_Toc211805451)

[Hình 1.1: Mô hình Client - Server 4](#_Toc211805452)

[Hình 1.2: AOF 5](#_Toc211805453)

[Hình 1.3: RDB 5](#_Toc211805454)

[Hình 1.4: Replication 7](#_Toc211805455)

[Hình 1.5: Cluster 9](#_Toc211805456)

[Hình 1.6: Sentinel 10](#_Toc211805457)

[Hình 1.7: Tài nguyên tải về 13](#_Toc211805458)

[Hình 1.8 Các tệp trong thư mục nén tải về 13](#_Toc211805459)

[Hình 1.11: Chạy lệnh ping kiểm tra kết nối 15](#_Toc211805460)

[Hình 1.12: Bật tính năng WSL trên Windows 16](#_Toc211805461)

[Hình 1.13: Cài đặt Ubuntu từ Microsoft Store 16](#_Toc211805462)

[Hình 1.14: Cài đặt 17](#_Toc211805463)

[Hình 1.15: Tải redis 17](#_Toc211805464)

[Hình 1.16: Chạy lệnh ping kiểm tra 18](#_Toc211805465)

[Hình 1.17: Cấu trúc dữ liệu 19](#_Toc211805466)

DANH MỤC CÁC KÝ HIỆU VÀ CHỮ VIẾT TẮT

|  |  |  |
| --- | --- | --- |
| **Viết tắt** | **Tiếng Anh** | **Tiếng Việt** |
| RDBMS | Relational Database Management System | Hệ quản trị cơ sở dữ liệu quan hệ |
| CRUD | Creat – Read – Update - Delete | Bốn thao tác cơ bản: Tạo - Đọc - Cập nhật - Xóa |
| RAM | Random Access Memory | Bộ nhớ truy cập ngẫu nhiên |
| CSDL |  | Cơ sở dữ liệu |
| WSL | Windows Subsystem for Linux |  |
| AOF | Append Only File |  |
| RDB | Redis Database |  |
|  |  |  |

# Tổng quan về Redis

## Giới thiệu Redis

### Redis là gì?

Redis (REmote DIctionary Server) là một hệ thống lưu trữ cấu trúc dữ liệu trong bộ nhớ (in-memory data structure store), mã nguồn mở. Redis có thể được sử dụng như một cơ sở dữ liệu, bộ nhớ đệm (cache), message broker, hoặc streaming engine.

Khác với cơ sở dữ liệu truyền thống lưu trữ trên đĩa, Redis lưu dữ liệu trong RAM, từ đó đạt được tốc độ xử lý cực nhanh và hiệu quả trong nhiều trường hợp sử dụng.

Hãy hình dung Redis giống như một kho key-value hiệu năng cao - tương tự một từ điển hoặc hash map khổng lồ - nhưng toàn bộ dữ liệu được lưu trong RAM. Chính nhờ thiết kế này mà Redis có thể xử lý hàng trăm nghìn lệnh mỗi giây với độ trễ rất thấp.

### Các tính năng nổi bật của Redis

Redis mang lại nhiều ưu điểm khiến nó trở nên phổ biến trong giới phát triển:

* Tốc độ cao: Hoạt động hoàn toàn trong bộ nhớ giúp Redis có độ trễ dưới 1ms.
* Đa dạng kiểu dữ liệu: Hỗ trợ nhiều cấu trúc như *string, hash, list, set, sorted set,...*
* Khả năng lưu bền (Persistence): Có thể lưu dữ liệu xuống đĩa định kỳ hoặc ghi log các thao tác để đảm bảo an toàn dữ liệu.
* Replication: Hỗ trợ mô hình master-slave (leader-follower replication) để tăng tính sẵn sàng.
* Clustering: Có thể triển khai kiến trúc phân tán.
* Lua scripting: Cho phép chạy script tuỳ chỉnh ngay trong Redis.
* Transactions: Hỗ trợ giao dịch cơ bản.
* Pub/Sub: Cung cấp mô hình publish/subscribe cho việc nhắn tin thời gian thực.

### Khi nào nên sử dụng Redis?

Redis đặc biệt phù hợp trong các tình huống yêu cầu tốc độ cao và xử lý tức thì:

* Caching: Lưu trữ dữ liệu truy cập thường xuyên để giảm tải cho database chính.
* Quản lý session: Lưu trữ session người dùng trong các ứng dụng web.
* Phân tích thời gian thực: Đếm và theo dõi các sự kiện ngay khi chúng xảy ra.
* Leaderboard/Counting: Sử dụng sorted set để xây dựng bảng xếp hạng hoặc bộ đếm.
* Hàng đợi (Queues): Dùng list để xử lý các tác vụ theo thứ tự.
* Giao tiếp thời gian thực: Triển khai chat, messaging với Pub/Sub.

### Redis so với Cơ sở dữ liệu truyền thống

**Sơ đồ so sánh**

A diagram of a diagram

AI-generated content may be incorrect.

*Giải thích sơ đồ:*

Một yêu cầu từ Client sẽ được gửi đến Database:

* Nếu cơ sở dữ liệu được chọn là Redis, dữ liệu sẽ được xử lý trực tiếp trong bộ nhớ (In-Memory Processing) và phản hồi lại rất nhanh.
* Nếu cơ sở dữ liệu được chọn là RDBMS truyền thống, dữ liệu sẽ được xử lý thông qua truy xuất ổ đĩa (Disk I/O) trước khi phản hồi, nên tốc độ chậm hơn.

**Redis có thay thế hoàn toàn RDBMS không?**

Redis không được thiết kế để thay thế toàn bộ cơ sở dữ liệu truyền thống trong mọi tình huống. Thay vào đó, mỗi hệ thống có ưu điểm riêng. Dưới đây là bảng so sánh:

Bảng 1.1: Bảng so sánh Redis với RDBMS

|  |  |  |
| --- | --- | --- |
| Tiêu chí | Redis | Cơ sở dữ liệu truyền thống (RDBMS) |
| Lưu trữ dữ liệu | Trong bộ nhớ (RAM), có thể cấu hình thêm lưu bền | Dựa trên ổ đĩa |
| Ngôn ngữ truy vấn | Lệnh đơn giản | SQL hoặc tương tự |
| Hỗ trợ cấu trúc dữ liệu | Nhiều cấu trúc dựng sẵn: string, hash, list, set, sorted set, … | Bảng với khả năng quan hệ |
| Hỗ trợ giao dịch | Cơ bản (Basic transaction) | Tuân thủ ACID đầy đủ |
| Yêu cầu bộ nhớ | Cao hơn (dữ liệu nằm trong RAM) | Thấp hơn (dùng ổ đĩa) |
| Tốc độ | Rất nhanh (microseconds) | Trung bình đến nhanh (milliseconds) |

## Kiến trúc

### Thành phần Client-Server

**Redis Server**: Là lõi của Redis, xử lý các yêu cầu, quản lý dữ liệu và đảm bảo tính bền vững. Redis là một kho khóa-giá trị trong bộ nhớ.

* Mô hình đơn luồng: Redis xử lý tất cả các lệnh tuần tự bằng một luồng duy nhất, giúp tránh sự phức tạp và chi phí của các cơ chế đồng bộ hóa đa luồng.
* Vòng lặp sự kiện: Server sử dụng kiến trúc hướng sự kiện với vòng lặp I/O không chặn (epoll, kqueue, select) để xử lý hàng ngàn kết nối đồng thời mà không bị trễ.

**Client Libraries**: Hỗ trợ nhiều ngôn ngữ, gồm có hai loại là client trực tiếp và client hỗ trợ Object Mapping.

**Hoạt động dữ liệu**: Redis hỗ trợ phong phú các kiểu dữ liệu giá trị (Strings, Lists, Sets, Hashes...) và các thao tác CRUD cơ bản.

A diagram of a server

AI-generated content may be incorrect.

Hình 1.1: Mô hình Client - Server

### Persistence (Tính bền vững)

Redis lưu trữ dữ liệu trong bộ nhớ để truy cập siêu nhanh, nhưng cũng có các tùy chọn lưu trữ trên đĩa để phục hồi dữ liệu. Có 4 tùy chọn chính:

**AOF (Append Only File):**

* Hoạt động như một nhật ký cam kết, ghi lại mọi thao tác ghi vào Redis.
* Có ba chế độ: **Always** (đồng bộ mọi lệnh, bền nhất nhưng chậm nhất), **Everysec** (đồng bộ mỗi giây, cân bằng), và **No** (để HĐH xử lý, nhanh nhất nhưng kém bền hơn).

A diagram of a server

AI-generated content may be incorrect.

Hình 1.2: AOF

* Cấu hình:

|  |
| --- |
| # Redis configuration for AOF  appendonly yes  appendfsync everysec # Sync AOF once per second |

**RDB (Redis Database):**

* Thực hiện các snapshot tại các khoảng thời gian được định trước.
* Snapshot được tạo bằng cách sử dụng một tiến trình con (forked process). Tiến trình cha tiếp tục phục vụ các client, trong khi tiến trình con ghi dữ liệu vào đĩa.
* **Ưu điểm** là các file nhỏ gọn và ít tác động đến hiệu suất.
* **Nhược điểm** là có nguy cơ mất dữ liệu giữa các snapshot.

A close-up of a logo

AI-generated content may be incorrect.

Hình 1.3: RDB

* Cấu hình:

|  |
| --- |
| # Redis configuration for RDB  save 900 1 # Create a snapshot if at least 1 key changes in 900 seconds  save 300 10 # Create a snapshot if at least 10 keys change in 300 seconds  save 60 10000 # Create a snapshot if at least 10000 keys change in 60 seconds |

**AOF và RDB (Hybrid Persistence):**

* Phương pháp này kết hợp ưu điểm của cả AOF và RDB.
* Trong quá trình phục hồi, Redis tải snapshot RDB trước và sau đó phát lại các log AOF gần đây, giúp tăng tốc độ phục hồi.
* Cấu hình:

|  |
| --- |
| # Redis using both persistence methods  appendonly yes  save 900 1 |

**No Persistence:**

* Thường được sử dụng khi Redis đóng vai trò là cache cho các tập dữ liệu nhỏ hoặc dữ liệu ít quan trọng, nơi việc mất dữ liệu không phải là vấn đề quá lớn. Khi Redis sập, toàn bộ dữ liệu sẽ mất.

### Replication (Nhân bản)

Redis áp dụng cơ chế replication theo mô hình master–slave (leader–follower) để đảm bảo tính sẵn sàng cao:

* Master (Leader): Xử lý tất cả các thao tác ghi và truyền các thay đổi dữ liệu đến các bản sao.
* Slaves (Followers): Hoạt động như bản sao chỉ đọc của master, hỗ trợ failover và có thể xử lý nhiều yêu cầu đọc song song. Theo mặc định, các bản sao từ chối lệnh ghi để giữ tính nhất quán dữ liệu.

Quá trình replication diễn ra không đồng bộ (asynchronous), nhờ đó master vẫn duy trì hiệu suất cao và độ trễ thấp. Các bản sao sẽ định kỳ xác nhận dữ liệu đã nhận được.

Mỗi master được gắn với Replication ID và offset (tự động tăng theo từng thao tác) để theo dõi trạng thái đồng bộ.

Redis hỗ trợ hai hình thức đồng bộ:

* Đồng bộ toàn bộ (Full Sync): Khi bản sao không có offset phù hợp hoặc không biết offset, master sẽ tạo một RDB snapshot và gửi sang. Trong quá trình này, master lưu tạm (buffer) các lệnh ghi để đồng bộ sau.
* Đồng bộ một phần (Partial Sync): Khi master và replica chia sẻ cùng một ancestor ID, hệ thống chỉ truyền các phần dữ liệu chênh lệch dựa trên offset, giúp tiết kiệm thời gian và băng thông.

A diagram of a diagram

AI-generated content may be incorrect.

Hình 1.4: Replication

Cấu hình:

|  |
| --- |
| # On the replica server  replicaof 192.168.1.100 6379 |

### Cluster

Redis Cluster cho phép mở rộng theo chiều ngang bằng cách phân vùng dữ liệu trên nhiều node, dựa trên cơ chế master–slave replication.

Một cluster bao gồm nhiều shard, trong đó mỗi shard có một node primary và có thể kèm theo nhiều node replica.

**Các đặc điểm chính:**

* Sharding: Dữ liệu được tự động phân vùng và phân phối trên các node primary khi số lượng shard tăng.
* Hash Slots: Redis Cluster chia dữ liệu thành 16384 hash slots. Khóa được ánh xạ vào slot bằng hàm CRC16(key) mod 16384, và mỗi node chịu trách nhiệm cho một nhóm slot.
* Hash Tags: Hỗ trợ thao tác đa khóa bằng cách sử dụng dấu ngoặc nhọn {} để đảm bảo các khóa liên quan nằm trong cùng một slot.
* Resharding: Khi thêm hoặc bớt shard, hệ thống di chuyển các hash slot thay vì dữ liệu trực tiếp, giúp giảm chi phí và hạn chế ảnh hưởng đến tính sẵn sàng.
* Failover: Nếu một master node gặp sự cố, một replica sẽ được thăng cấp thành master mới.
* Gossiping: Sử dụng giao thức gossip để giám sát tình trạng cluster và thay thế node khi cần.
* Sharded Pub/Sub (Redis 7.0): Cho phép Pub/Sub theo shard, chỉ gửi tin nhắn trong shard cụ thể, hiệu quả hơn so với Pub/Sub toàn cluster.

**Ưu điểm:**

* Phân vùng dữ liệu tự động.
* Mở rộng cả đọc và ghi.
* Đảm bảo hoạt động liên tục ngay cả khi một phần node gặp lỗi.

**Nhược điểm:**

* Cần client tương thích với Redis Cluster.
* Thiết kế ứng dụng phải tính đến ràng buộc thao tác đa khóa.
* Kiến trúc phức tạp hơn khi vận hành ở quy mô lớn.

A diagram of a computer program

AI-generated content may be incorrect.

Hình 1.5: Cluster

**Cấu hình:**

|  |
| --- |
| # Creating a simple 3-node cluster  redis-cli --cluster create 127.0.0.1:7000 127.0.0.1:7001 127.0.0.1:7002 |

### Sentinel

Redis Sentinel là một hệ thống phân tán được thiết kế để quản lý và giám sát Redis, với nhiều tiến trình Sentinel hoạt động phối hợp cùng nhau.

**Chức năng chính:**

* Tính sẵn sàng cao: Liên tục kiểm tra tình trạng của các instance primary và replica.
* Tự động Failover: Khi primary gặp sự cố, Sentinel sẽ thăng cấp một replica thành primary mới và cấu hình lại các replica khác để theo dõi node mới này.
* Quorum-based: Các quyết định failover được thông qua bởi đa số Sentinel, ngăn chặn tình huống split-brain.
* Service Discovery: Client có thể hỏi Sentinel để biết địa chỉ primary hiện tại. Khi failover xảy ra, Sentinel sẽ cung cấp thông tin cập nhật.
* Thông báo: Có khả năng gửi cảnh báo cho quản trị viên hoặc hệ thống liên quan về các sự kiện quan trọng.

**Ưu điểm:**

* Tự động failover khi có sự cố.
* Hỗ trợ service discovery cho client.

**Nhược điểm:**

* Kiến trúc phức tạp hơn vì cần duy trì một nhóm Sentinel riêng.
* Client phải hỗ trợ Sentinel.

A diagram of a company

AI-generated content may be incorrect.

Hình 1.6: Sentinel

**Cấu hình:**

|  |
| --- |
| # sentinel.conf  sentinel monitor mymaster 127.0.0.1 6379 2  sentinel down-after-milliseconds mymaster 5000  sentinel failover-timeout mymaster 60000 |

### Memory Management (Quản lý bộ nhớ)

Redis lưu trữ toàn bộ dữ liệu trong RAM để đạt tốc độ truy cập cực nhanh, vì vậy việc quản lý bộ nhớ hiệu quả là yếu tố then chốt.

**Cấu trúc dữ liệu trong bộ nhớ**

* Hash Table: Lưu trữ tất cả cặp key–value, tra cứu O(1). Khi xảy ra xung đột hash, Redis dùng phương pháp chaining và thực hiện rehash bằng hai bảng hash toàn cục khi cần.
* SDS (Simple Dynamic Strings): Chuỗi động được cải tiến từ C-string, hỗ trợ cả dữ liệu nhị phân. Có thuộc tính len (độ dài) và free (vùng trống cấp phát sẵn), giúp thao tác chuỗi O(1) và an toàn bộ nhớ.
* Ziplist: Cấu trúc nhỏ gọn cho list/hash nhỏ, dữ liệu lưu liên tục trong bộ nhớ. Redis 7.0 thay Ziplist bằng Listpack và dùng Quicklist cho list lớn.
* Intset: Cấu trúc tối ưu cho các tập hợp số nguyên nhỏ.
* Quicklist: Kết hợp linked list và ziplist, được dùng cho các danh sách lớn.

**Chính sách trục xuất (Eviction Policies)**

Khi đạt giới hạn bộ nhớ, Redis áp dụng chiến lược trục xuất để giải phóng không gian (cấu hình ví dụ):

|  |
| --- |
| maxmemory 1gb  maxmemory-policy allkeys-lru |

Các chính sách bao gồm:

* Noeviction: Từ chối ghi thêm dữ liệu mới.
* Volatile-lru: Xóa key ít dùng gần đây, chỉ trong tập có TTL.
* Allkeys-lru: Xóa key ít dùng gần đây, áp dụng cho toàn bộ key.
* Volatile-ttl: Xóa key sắp hết hạn trước.
* Allkeys-random: Xóa key ngẫu nhiên.

**Phân mảnh bộ nhớ**

Redis mặc định dùng Jemalloc để quản lý cấp phát và giảm phân mảnh bộ nhớ.

**Hết hạn dữ liệu & cơ chế xóa**

* TTL (Time To Live): Key có thể cài đặt thời gian sống.
* Lazy Deletion: Key chỉ bị xóa khi được truy cập lại sau khi hết hạn → tiết kiệm CPU.
* Active Expiration: Tiến trình nền định kỳ quét và xóa key hết hạn để giải phóng bộ nhớ kịp thời.

## Hướng dẫn cài đặt

### Cài đặt Redis trên Windows

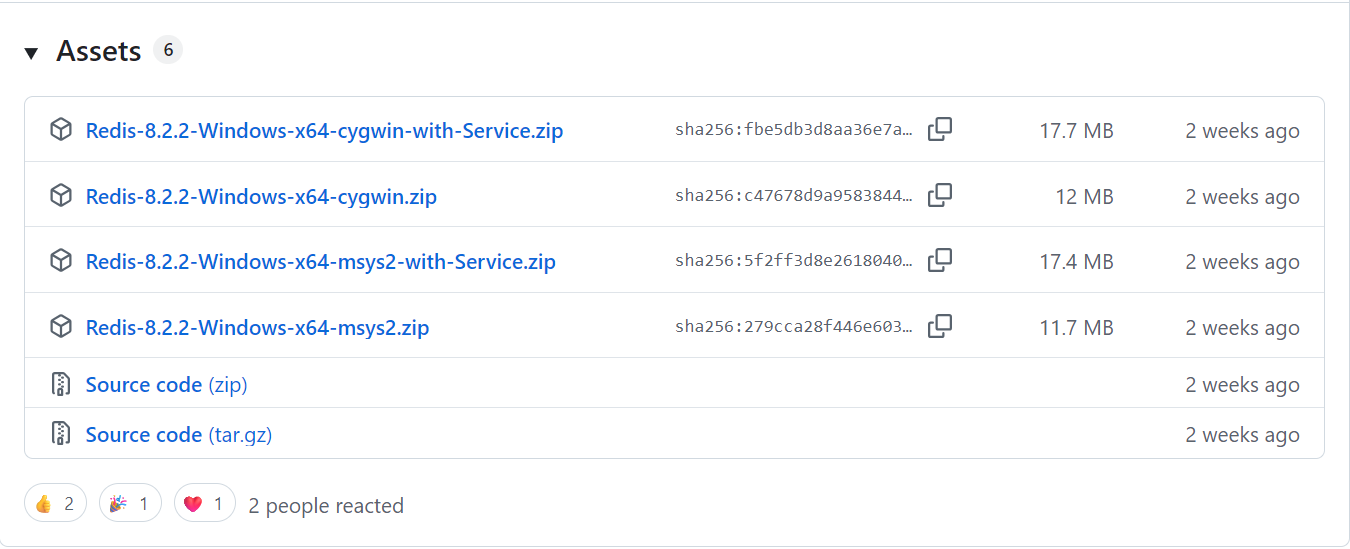
**Bước 1**: Truy cập vào địa chỉ: <https://github.com/redis-windows/redis-windows/releases> và tải xuống tệp cài đặt phù hợp. Tại đây, Redis phiên bản **8.2.2** được cung cấp dưới dạng các bản dựng (build) dành cho môi trường Windows, với hai tùy chọn chính:

* **Cygwin build:** thích hợp cho người dùng cần môi trường tương thích với Unix, tuy nhiên dung lượng lớn hơn và yêu cầu nhiều tài nguyên hơn.
* **MSYS2 build:** gọn nhẹ, hoạt động tối ưu trong môi trường Windows thông thường.

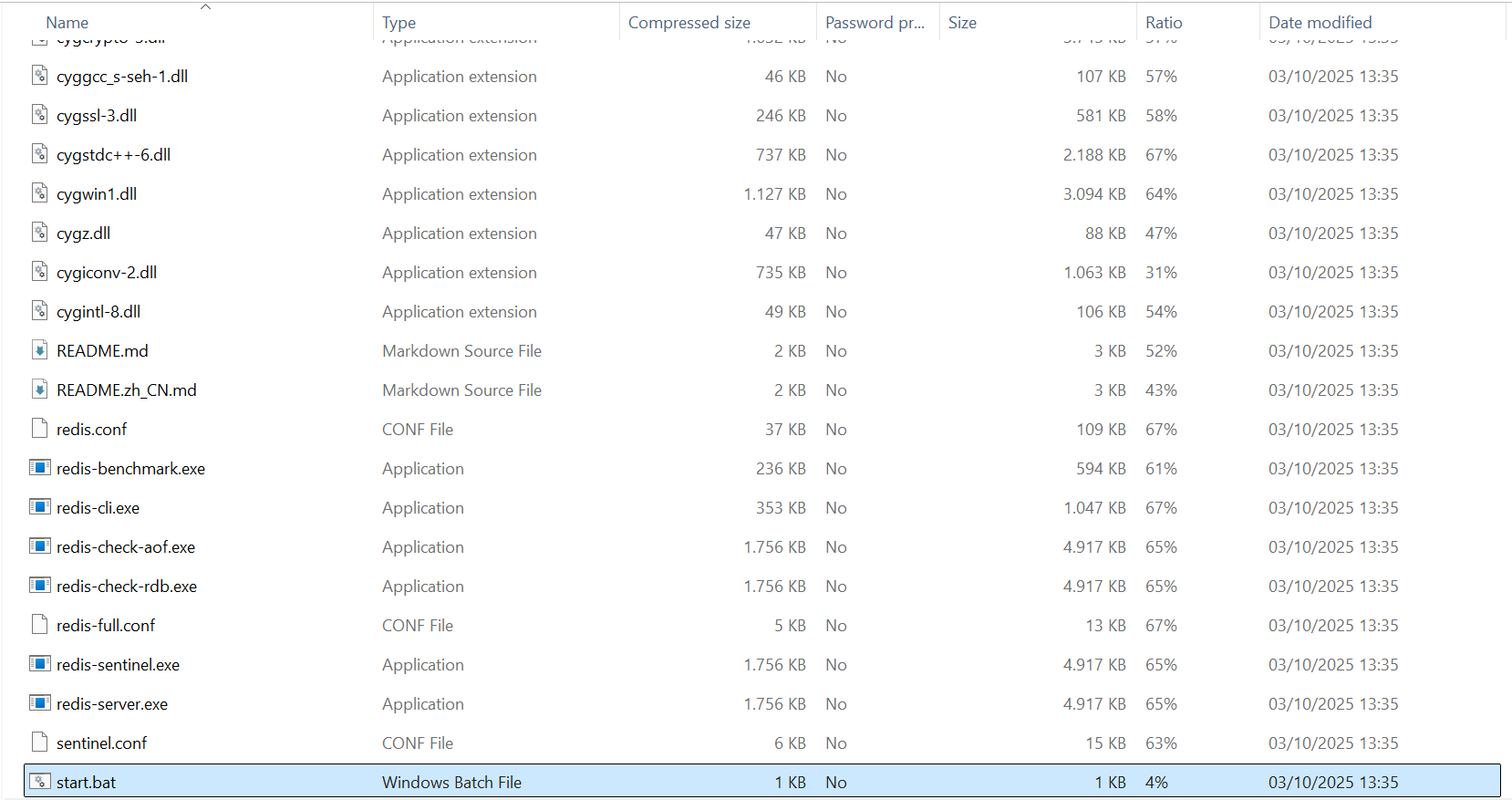
Mỗi tùy chọn đều có hai biến thể:

* **Bản “with-Service”** (ví dụ: Redis-8.2.2-Windows-x64-cygwin-with-Service.zip, dung lượng khoảng 17.7 MB), cho phép Redis tự động khởi động và hoạt động như một **dịch vụ hệ thống**.
* **Bản không “with-Service”** (ví dụ: Redis-8.2.2-Windows-x64-cygwin.zip, khoảng 12 MB), thích hợp khi người dùng muốn **chạy Redis thủ công**.

Đối với hầu hết người dùng Windows, **bản MSYS2** là lựa chọn khuyến nghị. Trong khi đó, **bản Cygwin** phù hợp hơn với các nhà phát triển cần môi trường gần với hệ thống Unix. Ngoài ra, **mã nguồn Redis** (định dạng .zip hoặc .tar.gz) cũng được cung cấp để người dùng tùy chỉnh hoặc biên dịch theo nhu cầu riêng.



Hình 1.7: Tài nguyên tải về



Hình 1.8 Các tệp trong thư mục nén tải về

Dự án Redis dành cho Windows cung cấp ba phương thức chính để khởi động và vận hành hệ thống Redis, phù hợp với các mục đích và mức độ sử dụng khác nhau. Cụ thể như sau:

1. **Khởi chạy trực tiếp bằng tệp thực thi (*start.bat*)**

Đây là phương thức khởi động nhanh và thuận tiện nhất. Người dùng chỉ cần nhấp đúp vào tệp start.bat trong thư mục cài đặt để Redis được khởi động ngay lập tức mà không cần thao tác qua dòng lệnh.

Phương pháp này đặc biệt phù hợp cho môi trường thử nghiệm hoặc mục đích học tập.

1. Khởi chạy Redis thông qua dòng lệnh

Redis có thể được khởi động trực tiếp bằng *Command Prompt* hoặc *PowerShell* với tệp cấu hình redis.conf như sau:

* **Command Prompt (CMD):** redis-server.exe redis.conf
* **PowerShell:** ./redis-server.exe redis.conf

Phương pháp này cho phép người dùng dễ dàng tùy chỉnh các tham số cấu hình trong tệp redis.conf, phù hợp cho các môi trường phát triển cần kiểm soát chi tiết.

1. Cài đặt và vận hành Redis như một dịch vụ hệ thống Windows

Phương pháp này giúp Redis có thể tự động khởi động cùng hệ thống và hoạt động ổn định ở chế độ nền (background service).

Quá trình cài đặt và quản lý dịch vụ Redis được thực hiện qua các lệnh sau:

* **Cài đặt dịch vụ:**

|  |
| --- |
| sc.exe create Redis binpath=C:\Software\Redis\RedisService.exe start= auto |

* **Khởi động dịch vụ:**

|  |
| --- |
| net start Redis |

* **Dừng dịch vụ:**

|  |
| --- |
| net stop Redis |

* **Gỡ bỏ dịch vụ khỏi hệ thống:**

|  |
| --- |
| sc.exe delete Redis |

Việc chạy Redis như một dịch vụ hệ thống đặc biệt hữu ích trong các môi trường sản xuất hoặc hệ thống cần đảm bảo tính sẵn sàng cao, giúp Redis tự động khởi động khi máy chủ khởi động lại.

**Bước 2:** Kết nối với Redis

Mở **Terminal/Command Prompt**.

Gõ lệnh:

* **redis-cli:** mở Redis Command Line Interface.
* **ping***:* kiểm tra kết nối giữa client và server Redis.

Nếu kết quả trả về là **PONG**, nghĩa là Redis đã được cài đặt thành công và đang hoạt động bình thường trên máy.

A screenshot of a computer

AI-generated content may be incorrect.

Hình 1.11: Chạy lệnh ping kiểm tra kết nối

### Cài đặt Redis thông qua WSL (Windows Subsystem for Linux)

Ngoài việc cài trực tiếp trên Windows, bạn cũng có thể chạy Redis thông qua môi trường Linux trên Windows với WSL. Cách thực hiện như sau:

Mở PowerShell (với quyền Administrator) và nhập “*Enable-WindowsOptionalFeature -Online -FeatureName Microsoft-Windows-Subsystem-Linux*”.

A screenshot of a computer

AI-generated content may be incorrect.

Hình 1.12: Bật tính năng WSL trên Windows

Sau khi WSL được kích hoạt, truy cập **Microsoft Store**, tìm kiếm và cài đặt bản phân phối **Ubuntu**.

A screenshot of a computer

AI-generated content may be incorrect.

Hình 1.13: Cài đặt Ubuntu từ Microsoft Store

Mở ứng dụng **Ubuntu** và tiến hành cài đặt Redis theo sự hướng dẫn của trang [Install Redis on Windows | Docs](https://redis.io/docs/latest/operate/oss_and_stack/install/archive/install-redis/install-redis-on-windows/).

A screenshot of a computer

AI-generated content may be incorrect.

Hình 1.14: Cài đặt

A screenshot of a computer screen

AI-generated content may be incorrect.

Hình 1.15: Tải redis

Kết nối với Redis.

A screenshot of a computer

AI-generated content may be incorrect.

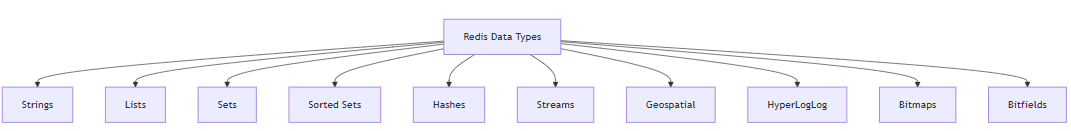
Hình 1.16: Chạy lệnh ping kiểm tra

### Các công cụ GUI tiêu biểu cho Redis

Redis cung cấp nhiều công cụ giao diện trực quan, giúp người dùng dễ dàng quản lý và tương tác với dữ liệu. Dưới đây là bốn công cụ phổ biến nhất: RedisInsight, Redis Desktop Manager (RDM), P3X Redis UI và Medis.

* **RedisInsight**: Được phát triển bởi Redis Labs, RedisInsight nổi bật với giao diện thân thiện, hỗ trợ trực quan hóa dữ liệu, phân tích Slowlog và giám sát hiệu suất. Đây là lựa chọn toàn diện cho các nhà phát triển, mặc dù một số tính năng nâng cao yêu cầu đăng ký trả phí.
* **Redis Desktop Manager (RDM)**: Là một trong những công cụ được ưa chuộng nhất, RDM hỗ trợ hiển thị cấu trúc dữ liệu rõ ràng, bảo mật kết nối qua SSL/TLS và SSH tunneling, đồng thời tích hợp giao diện dòng lệnh (CLI) để thực thi lệnh. Với sự ổn định và dễ sử dụng, RDM được cộng đồng lập trình viên đánh giá cao.
* **P3X Redis UI**: Là công cụ mã nguồn mở dựa trên Electron, P3X Redis UI hỗ trợ quản lý Redis Cluster và Sentinel. Công cụ này phù hợp cho người dùng cần sự linh hoạt, có thể hoạt động cả trực tuyến và ngoại tuyến, nhưng giao diện có thể hơi phức tạp với người mới bắt đầu.
* **Medis**: Tập trung vào sự đơn giản và gọn nhẹ, Medis cho phép người dùng dễ dàng xem và chỉnh sửa dữ liệu. Với thiết kế tối giản, công cụ này lý tưởng cho những ai cần thao tác nhanh chóng, mặc dù không có nhiều tính năng nâng cao.

## Cấu trúc dữ liệu trong Redis



Hình 1.17: Cấu trúc dữ liệu

### Strings

#### Giới thiệu

String là loại dữ liệu cơ bản nhất trong Redis, dùng để lưu trữ chuỗi văn bản, số, hoặc dữ liệu nhị phân (tối đa 512MB). String phù hợp cho các giá trị key-value đơn giản như cấu hình, bộ đếm, hoặc dữ liệu cache.

#### Lệnh cơ bản

|  |  |  |
| --- | --- | --- |
| Lệnh | Cú pháp | Mô tả |
| SET | SET key value [options] | Thiết lập giá trị cho key. |
| GET | GET key | Lấy giá trị của key. |
| DEL | DEL key [key ...] | Xóa một hoặc nhiều key. |
| INCR | INCR key | Tăng giá trị số nguyên lên 1 (khởi tạo 0 nếu chưa tồn tại). |
| DECR | DECR key | Giảm giá trị số nguyên xuống 1. |
| APPEND | APPEND key value | Thêm chuỗi vào giá trị hiện tại. |
| STRLEN | STRLEN key | Trả về độ dài chuỗi. |

#### Ví dụ minh họa

**Thiết lập và thao tác với String**

|  |
| --- |
| SET username “VuDang” *# Tạo key 'username' với giá trị “VuDang*”  GET username *# Lấy giá trị của key*  STRLEN username *# Trả về độ dài chuỗi -> 6*  GETRANGE username 0 1 *# Lấy ký tự từ vị trí [0,1]*  SETRANGE username 7 “Redis!” *# Thay thế từ vị trí thứ 7*  APPEND username “- HUIT!” *# Nối thêm vào cuối chuỗi* |

**Set/Get nhiều key cùng lúc**

|  |
| --- |
| MSET first\_name “Vu” last\_name “Dang” age 21 *# Đặt nhiều key cùng lúc*  MGET first\_name last\_name age *# Lấy nhiều key cùng lúc* |

**Key có thời gian sống (TTL)**

|  |
| --- |
| SET session:user123 “active” EX 3600 *# Tạo key tồn tại 3600 giây (1h)*  SETEX temp\_token 30 “a1b2c3d4” *# Viết gọn, key sống 30 giây*  TTL session:user123 *# Dùng TTL kiểm tra* |

**Set với điều kiện (NX, XX)**

|  |
| --- |
| SET counter 10 NX *# Chỉ set nếu key chưa tồn tại*  SET counter 20 XX *# Chỉ set nếu key đã tồn tại* |

**Tăng giảm giá trị số**

|  |
| --- |
| SET counter 10  INCR counter *# +1 => 11*  GET counter *# -> “11”*  INCRBY counter 5 *# +5 => 16*  GET counter *# -> “16”* |

#### Ứng dụng thực tế

* Caching dữ liệu web: Lưu trữ kết quả truy vấn cơ sở dữ liệu (như danh sách sản phẩm) để giảm tải server chính. Ví dụ, một trang thương mại điện tử có thể lưu trữ nội dung trang chủ trong Redis để trả về nhanh chóng cho người dùng.
* Bộ đếm thời gian thực: Theo dõi số lượt xem bài viết, lượt thích, hoặc số lượng người dùng trực tuyến trên một website. Ví dụ, sử dụng INCR để đếm số lượt truy cập vào một bài blog.
* Cấu hình ứng dụng: Lưu trữ các giá trị cấu hình động, như URL API hoặc thông số hệ thống, để dễ dàng cập nhật mà không cần khởi động lại ứng dụng.
* Lưu trữ token phiên: Lưu trữ các token xác thực tạm thời (như JWT) với thời gian sống (TTL) để quản lý phiên đăng nhập người dùng.

### Lists

#### Giới thiệu

List là danh sách có thứ tự, cho phép thêm/xóa phần tử ở đầu hoặc cuối, hoạt động như hàng đợi (queue) hoặc ngăn xếp (stack).

#### Lệnh cơ bản

|  |  |  |
| --- | --- | --- |
| Lệnh | Cú pháp | Mô tả |
| LPUSH | LPUSH key element [...] | Thêm phần tử vào đầu danh sách. |
| RPUSH | RPUSH key element [...] | Thêm phần tử vào cuối danh sách. |
| LPOP | LPOP key | Xóa và trả phần tử đầu. |
| RPOP | RPOP key | Xóa và trả phần tử cuối. |
| DEL | DEL key | Xóa danh sách nếu rỗng sau thao tác. |
| LRANGE | LRANGE key start stop | Trả về phạm vi phần tử (0 - 1 cho tất cả). |
| LLEN | LLEN key | Trả về độ dài danh sách. |

#### Ví dụ minh họa

|  |
| --- |
| # Thêm phần tử vào danh sách  LPUSH mylist “world” *# -> (integer) 1*  LPUSH mylist “hello” *# -> (integer) 2*  RPUSH mylist “Redis” *# -> (integer) 3*  LRANGE mylist 0 -1 *# Xem toàn bộ danh sách*  LINDEX mylist 0 *# Lấy phần tử đầu tiên (index = 0)*  LINDEX mylist -1 *# Lấy phần tử cuối cùng (index = -1)*  LPOP mylist *# Xóa và lấy phần tử đầu tiên*  RPOP mylist *# Xóa và lấy phần tử cuối cùng*  LRANGE mylist 0 -1 *# Xem lại danh sách còn gì* |

#### Ứng dụng thực tế

* Hàng đợi tác vụ bất đồng bộ: Quản lý các tác vụ như gửi email hàng loạt, xử lý thanh toán, hoặc tạo báo cáo. Ví dụ, một hệ thống gửi thông báo đẩy có thể sử dụng LPUSH để thêm thông báo và RPOP để xử lý.
* Lịch sử hành động người dùng: Lưu trữ danh sách các hành động gần đây, như lịch sử tìm kiếm hoặc sản phẩm đã xem trên website thương mại điện tử.
* Hàng đợi tin nhắn thời gian thực: Xử lý tin nhắn trong các ứng dụng chat, nơi tin nhắn được thêm vào danh sách bằng RPUSH và đọc bằng LRANGE.
* Hàng đợi xử lý sự kiện: Quản lý luồng sự kiện trong hệ thống IoT, như lưu trữ các tín hiệu từ thiết bị thông minh để xử lý theo thứ tự.

### Sets

#### Giới thiệu

Set là tập hợp không thứ tự, không cho phép trùng lặp, phù hợp để lưu trữ danh sách các mục duy nhất.

#### Lệnh cơ bản

|  |  |  |
| --- | --- | --- |
| Lệnh | Cú pháp | Mô tả |
| SADD | SADD key member [...] | Thêm một hoặc nhiều thành viên. |
| SMEMBERS | SMEMBERS key | Trả về tất cả thành viên. |
| SREM | SREM key member [...] | Xóa một hoặc nhiều thành viên. |
| DEL | DEL key | Xóa set nếu rỗng sau thao tác. |
| SCARD | SCARD key | Trả về số lượng thành viên. |
| SISMEMBER | SISMEMBER key member | Kiểm tra thành viên có tồn tại. |

#### Ví dụ minh họa

|  |
| --- |
| *# Thêm phần tử vào Set (chỉ giữ unique, không trùng)*  SADD myset “apple” “banana” “orange” “apple”  SMEMBERS myset *# Xem tất cả phần tử trong Set*  SISMEMBER myset “apple” *# Kiểm tra xem phần tử có trong Set không*  SISMEMBER myset “grape”  SREM myset “orange” *# Xóa phần tử*  SMEMBERS myset  SCARD myset *# Đếm số phần tử*  *# Thao tác Set nâng cao*  SADD set1 “a” “b” “c”  SADD set2 “b” “c” “d”  *# Giao nhau*  SINTER set1 set2  *# Hợp*  SUNION set1 set2  *# Khác biệt*  SDIFF set1 set2  SDIFF set2 set1 |

#### Ứng dụng thực tế

* Danh sách duy nhất: Lưu trữ danh sách người dùng trực tuyến, thẻ tag bài viết, hoặc danh mục sản phẩm để đảm bảo không trùng lặp. Ví dụ, một blog có thể dùng Set để lưu trữ các tag như “công nghệ”, “khoa học”.
* Xử lý giao dịch tập hợp: Sử dụng SINTER, SUNION, hoặc SDIFF để tìm bạn chung giữa hai người dùng trên mạng xã hội hoặc sản phẩm chung trong giỏ hàng của nhiều khách hàng.
* Kiểm tra thành viên nhanh: Xác minh xem một người dùng có trong danh sách đen (blacklist) hay không, như trong hệ thống kiểm duyệt nội dung.
* Quản lý quyền truy cập: Lưu trữ danh sách các vai trò hoặc quyền của người dùng, ví dụ, danh sách người dùng có quyền admin trong một ứng dụng.

### Hashes

#### Giới thiệu

Hash là cấu trúc dữ liệu bản đồ, lưu trữ các cặp trường-giá trị, phù hợp để biểu diễn các đối tượng như thông tin người dùng.

#### Lệnh cơ bản

|  |  |  |
| --- | --- | --- |
| Lệnh | Cú pháp | Mô tả |
| HSET | HSET key field value [...] | Thiết lập giá trị cho trường. |
| HGET | HGET key field | Lấy giá trị của trường. |
| HDEL | HDEL key field [...] | Xóa một hoặc nhiều trường. |
| DEL | DEL key | Xóa hash nếu không còn trường. |
| HGETALL | HGETALL key | Trả về tất cả trường và giá trị. |
| HLEN | HLEN key | Trả về số lượng trường. |

#### Ví dụ minh họa

|  |
| --- |
| *# Tạo một user với nhiều thuộc tính*  HSET user:1 username “vu” email “abc@gmail.com” age “21”  HGET user:1 username *# Lấy một trường cụ thể*  HMGET user:1 username email *# Lấy nhiều trường*  HGETALL user:1 *# Lấy toàn bộ fields và values*  HSET user:1 age “20” *# Cập nhật một trường*  HEXISTS user:1 email *# Kiểm tra field có tồn tại không*  HDEL user:1 age *# Xóa một trường*  HLEN user:1 *# Đếm số field còn lại* |

#### Ứng dụng thực tế

* Lưu trữ thông tin đối tượng: Biểu diễn thông tin người dùng (tên, tuổi, email) hoặc sản phẩm (tên, giá, mô tả) trong các ứng dụng như mạng xã hội hoặc thương mại điện tử.
* Cấu hình phức tạp: Lưu trữ các cấu hình có nhiều thuộc tính, như cài đặt giao diện người dùng hoặc thông số hệ thống, cho phép cập nhật từng trường riêng lẻ.
* Quản lý phiên người dùng: Lưu trữ dữ liệu phiên (session) như thông tin giỏ hàng hoặc trạng thái đăng nhập, với khả năng truy xuất nhanh các trường cụ thể.
* Lưu trữ dữ liệu cấu trúc: Dùng trong các hệ thống phân tích để lưu trữ các thuộc tính của sự kiện, như thông tin giao dịch (mã giao dịch, thời gian, giá trị).

### Sorted Sets

#### Giới thiệu

Sorted Set là tập hợp có thứ tự, mỗi thành viên có điểm số để sắp xếp, phù hợp cho bảng xếp hạng hoặc dữ liệu ưu tiên.

#### Lệnh cơ bản

|  |  |  |
| --- | --- | --- |
| Lệnh | Cú pháp | Mô tả |
| ZADD | ZADD key score member [...] | Thêm thành viên với điểm số. |
| ZRANGE | ZRANGE key start stop | Trả về phạm vi thành viên theo thứ tự. |
| ZREM | ZREM key member [...] | Xóa thành viên. |
| DEL | DEL key | Xóa sorted set. |
| ZCARD | ZCARD key | Trả về số lượng thành viên. |

#### Ví dụ minh họa

|  |
| --- |
| *# Thêm 3 người chơi với điểm*  ZADD leaderboard 100 “Alice”  ZADD leaderboard 200 “Bob”  ZADD leaderboard 150 “Charlie”  *# Xem toàn bộ bảng xếp hạng theo thứ tự tăng*  ZRANGE leaderboard 0 -1 WITHSCORES  *# Xem top 2 theo thứ tự giảm (người điểm cao nhất đứng đầu)*  ZREVRANGE leaderboard 0 1 WITHSCORES  *# Lấy hạng của Alice (bắt đầu từ 0)*  ZRANK leaderboard “Alice”  *# Lấy điểm của Bob*  ZSCORE leaderboard “Bob” |

#### Ứng dụng thực tế

* Bảng xếp hạng thời gian thực: Quản lý bảng xếp hạng người chơi trong trò chơi trực tuyến hoặc bài đăng phổ biến trên mạng xã hội dựa trên số lượt thích. Ví dụ, một ứng dụng game có thể dùng Sorted Set để xếp hạng người chơi theo điểm số.
* Hàng đợi ưu tiên: Quản lý các tác vụ theo mức độ ưu tiên, như xử lý đơn hàng VIP trước trong hệ thống thương mại điện tử.
* Tìm kiếm theo phạm vi: Lấy các mục trong khoảng điểm số, như tìm kiếm các bài viết có số lượt xem từ 100 đến 1000.
* Lập lịch tác vụ: Sắp xếp các tác vụ theo thời gian thực thi, ví dụ, trong hệ thống gửi thông báo định kỳ, với điểm số là timestamp.

### Bitmap

#### Giới thiệu

Bitmap là một chuỗi bit (0 hoặc 1) được sử dụng để lưu trữ và thao tác dữ liệu nhị phân một cách hiệu quả. Mỗi bit trong Bitmap đại diện cho một trạng thái, giúp tiết kiệm bộ nhớ khi xử lý tập hợp lớn các giá trị boolean

#### Lệnh cơ bản

|  |  |  |
| --- | --- | --- |
| Chỉ thị | Cú pháp / Giá trị mặc định | Mô tả |
| SETBIT | SETBIT key offset value | Đặt giá trị bit (0 hoặc 1) tại vị trí offset. |
| GETBIT | GETBIT key offset | Lấy giá trị bit tại vị trí offset. |
| BITCOUNT | BITCOUNT key [start end] | Đếm số bit 1 trong phạm vi (hoặc toàn bộ). |
| BITOP | BITOP operation destkey key [key ...] | Thực hiện phép toán bit (AND, OR, XOR, NOT) trên các key. |

#### Ví dụ minh họa

|  |
| --- |
| # Đặt bit tại vị trí 100 là 1 (đánh dấu người dùng đã xem)  SETBIT user:123:views 100 1  # Kiểm tra giá trị bit tại vị trí 100  GETBIT user:123:views 100  # Đặt thêm bit tại vị trí 101 là 1  SETBIT user:123:views 101 1  # Đếm tổng số bit 1 trong phạm vi 0-200  BITCOUNT user:123:views 0 200  # Thực hiện phép AND với một Bitmap khác  BITOP AND result user:123:views user:124:views  # Đặt bit tại vị trí 102 là 0 (đánh dấu không hoạt động)  SETBIT user:123:views 102 0  # Đếm lại số bit 1 sau thay đổi  BITCOUNT user:123:views  # Thực hiện phép OR với Bitmap khác  BITOP OR combined user:123:views user:125:views |

#### Ứng dụng thực tế

* Theo dõi trạng thái người dùng: Đánh dấu người dùng đã đăng nhập hoặc xem bài viết.
* Phân tích hành vi: Đếm ngày hoạt động liên tục của người dùng.
* Lọc dữ liệu: Sử dụng để lọc danh sách dựa trên trạng thái nhị phân.
* Quản lý quyền truy cập: Lưu trữ trạng thái quyền hạn của người dùng.

### HyperLogLog

#### Giới thiệu

HyperLogLog là một cấu trúc dữ liệu xác suất, được thiết kế để ước lượng số lượng phần tử duy nhất (cardinality) trong một tập hợp lớn với độ chính xác cao và sử dụng bộ nhớ rất nhỏ (thường chỉ khoảng 12KB).

#### Lệnh cơ bản

|  |  |  |
| --- | --- | --- |
| Chỉ thị | Cú pháp / Giá trị mặc định | Mô tả |
| PFADD | PFADD key element [element ...] | Thêm phần tử vào HyperLogLog. |
| PFCOUNT | PFCOUNT key [key ...] | Ước lượng số lượng phần tử duy nhất. |
| PFMERGE | PFMERGE destkey sourcekey [sourcekey ...] | Gộp nhiều HyperLogLog thành một. |

#### Ví dụ minh họa

|  |
| --- |
| # Thêm 3 lượt truy cập duy nhất  PFADD page:visits user1 user2 user3  # Ước lượng số lượng truy cập duy nhất  PFCOUNT page:visits  # Thêm thêm 2 lượt truy cập mới  PFADD page:visits user4 user5  # Gộp dữ liệu từ hai trang thăm  PFADD page:visits2 user6 user7 user8  PFMERGE total page:visits page:visits2  # Kiểm tra tổng số duy nhất sau khi gộp  PFCOUNT total  # Thêm thêm một lượt truy cập vào trang thứ nhất  PFADD page:visits user9  # Kiểm tra lại số duy nhất trên page:visits  PFCOUNT page:visits |

#### Ứng dụng thực tế

* Đếm người dùng truy cập duy nhất: Theo dõi số lượng người truy cập website.
* Phân tích dữ liệu lớn: Ước lượng số IP hoặc sự kiện không trùng lặp.
* Thống kê thời gian thực: Đếm số lượt xem video duy nhất.
* Kết hợp tập hợp: Gộp dữ liệu từ nhiều nguồn để phân tích tổng hợp.

### Streams

#### Giới thiệu

Streams là một kiểu dữ liệu dành cho việc lưu trữ và xử lý dữ liệu dạng luồng theo thứ tự thời gian, hoạt động như một nhật ký (log) có khả năng xử lý dữ liệu thời gian thực. Streams hỗ trợ thêm dữ liệu, đọc theo phạm vi, và quản lý nhóm người tiêu dùng (consumer groups) để xử lý phân tán.

#### Lệnh cơ bản

|  |  |  |
| --- | --- | --- |
| Chỉ thị | Cú pháp / Giá trị mặc định | Mô tả |
| XADD | XADD key \* field value [field value ...] | Thêm entry vào stream. |
| XRANGE | XRANGE key start end | Trả về phạm vi entry theo ID. |
| XREAD | XREAD STREAMS key [key ...] ID [ID ...] | Đọc entry từ stream (có thể block). |
| XDEL | XDEL key ID [ID ...] | Xóa entry theo ID. |

#### Ví dụ minh họa

|  |
| --- |
| *# Thêm một sự kiện nhiệt độ từ cảm biến*  *XADD sensor:stream \* sensor\_id 001 temp 25.5*  *# Thêm thêm một sự kiện khác*  *XADD sensor:stream \* sensor\_id 002 temp 26.0*  *# Đọc tất cả các sự kiện từ đầu stream*  *XRANGE sensor:stream - +*  *# Đọc sự kiện mới nhất (chờ nếu không có)*  *XREAD STREAMS sensor:stream $*  *# Xóa sự kiện đầu tiên*  *XDEL sensor:stream 123456789*  *# Thêm sự kiện với nhiều trường*  *XADD sensor:stream \* sensor\_id 003 temp 27.0 humidity 60*  *# Đọc các sự kiện từ ID cụ thể đến cuối*  *XRANGE sensor:stream 123456789 +* |

#### Ứng dụng thực tế

* Quản lý log hệ thống: Lưu trữ log hoạt động server.
* Hàng đợi tin nhắn thời gian thực: Xử lý tin nhắn trong ứng dụng chat.
* Ghi nhận sự kiện IoT: Lưu trữ dữ liệu cảm biến từ thiết bị thông minh.
* Phân tích luồng dữ liệu: Theo dõi sự kiện liên tục như giao dịch tài chính.

### Geo

#### Giới thiệu

Geo cho phép lưu trữ và truy vấn dữ liệu địa lý, như tọa độ (kinh độ, vĩ độ), và hỗ trợ các phép tính như tìm khoảng cách giữa hai điểm hoặc tìm các điểm trong một bán kính cụ thể.

#### Lệnh cơ bản

|  |  |  |
| --- | --- | --- |
| Chỉ thị | Cú pháp / Giá trị mặc định | Mô tả |
| GEOADD | GEOADD key longitude latitude member | Thêm tọa độ địa lý. |
| GEODIST | GEODIST key member1 member2 [unit] | Tính khoảng cách giữa hai điểm. |
| GEORADIUS | GEORADIUS key longitude latitude radius unit | Tìm điểm trong bán kính. |

#### Ví dụ minh họa

|  |
| --- |
| *# Thêm tọa độ của hai cửa hàng*  *GEOADD stores 106.66 10.76 “StoreA” 106.70 10.80 “StoreB”*  *# Thêm tọa độ của cửa hàng thứ ba*  *GEOADD stores 106.65 10.75 “StoreC”*  *# Tính khoảng cách giữa StoreA và StoreB (đơn vị km)*  *GEODIST stores “StoreA” “StoreB” km*  *# Tìm cửa hàng trong bán kính 5km quanh tọa độ 106.66 10.76*  *GEORADIUS stores 106.66 10.76 5 km*  *# Thêm tọa độ của một nhà hàng*  *GEOADD stores 106.68 10.78 “RestaurantX”*  *# Tính khoảng cách từ StoreA đến RestaurantX*  *GEODIST stores “StoreA” “RestaurantX” km* |

#### Ứng dụng thực tế

* Tìm cửa hàng gần nhất: Gợi ý cửa hàng dựa trên vị trí người dùng.
* Theo dõi phương tiện: Theo dõi vị trí xe buýt hoặc xe giao hàng.
* Kết nối xã hội: Tìm bạn bè trong khu vực dựa trên tọa độ.
* Ứng dụng du lịch: Tìm điểm tham quan trong bán kính nhất định.

### JSON

#### Giới thiệu

RedisJSON (thông qua mô-đun RedisJSON) cho phép lưu trữ, truy vấn và thao tác trực tiếp trên dữ liệu dạng JSON, giúp quản lý các tài liệu phức tạp mà không cần chuyển đổi định dạng.

#### Lệnh cơ bản

|  |  |  |
| --- | --- | --- |
| Chỉ thị | Cú pháp / Giá trị mặc định | Mô tả |
| JSON.SET | JSON.SET key path value | Lưu trữ giá trị JSON tại đường dẫn. |
| JSON.GET | JSON.GET key [path] | Lấy giá trị JSON tại đường dẫn. |
| JSON.DEL | JSON.DEL key path | Xóa phần tử JSON tại đường dẫn. |

#### Ví dụ minh họa

|  |
| --- |
| *# Lưu thông tin người dùng dưới dạng JSON*  *JSON.SET user:123 . '{“name”: “John”, “age”: 30, “city”: “New York”}'*  *# Lấy tên của người dùng*  *JSON.GET user:123 .name*  *# Cập nhật tuổi thành 31*  *JSON.SET user:123 .age 31*  *# Thêm trường email*  *JSON.SET user:123 .email* [*john@example.com*](mailto:john@example.com)  *# Xóa trường city*  *JSON.DEL user:123 .city*  *# Thêm danh sách sở thích*  *JSON.SET user:123 .hobbies “[\”reading\”, \”gaming\”]”*  *# Lấy toàn bộ dữ liệu người dùng*  *JSON.GET user:123 .* |

#### Ứng dụng thực tế

* Lưu trữ thông tin người dùng: Quản lý dữ liệu người dùng phức tạp.
* Cấu hình ứng dụng: Lưu trữ cấu hình động dạng JSON.
* Dữ liệu sản phẩm: Lưu thông tin sản phẩm trong thương mại điện tử.
* API thời gian thực: Truy xuất nhanh dữ liệu JSON trong ứng dụng web.

### TimeSeries

#### Giới thiệu

RedisTimeSeries (thông qua mô-đun RedisTimeSeries) được tối ưu hóa để lưu trữ và phân tích dữ liệu chuỗi thời gian, như số liệu cảm biến, giá cổ phiếu, hoặc thống kê hiệu suất hệ thống. Nó hỗ trợ nén dữ liệu và truy vấn theo thời gian.

#### Lệnh cơ bản

|  |  |  |
| --- | --- | --- |
| Chỉ thị | Cú pháp / Giá trị mặc định | Mô tả |
| TS.ADD | TS.ADD key timestamp value | Thêm giá trị tại timestamp. |
| TS.RANGE | TS.RANGE key from to | Truy vấn dữ liệu trong khoảng thời gian. |
| TS.DEL | TS.DEL key from to | Xóa dữ liệu trong khoảng thời gian. |

#### Ví dụ minh họa

|  |
| --- |
| *# Thêm giá trị nhiệt độ tại thời điểm 1697059200 (Unix timestamp)*  *TS.ADD sensor:temp 1697059200 25.5*  *# Thêm thêm giá trị tại thời điểm khác*  *TS.ADD sensor:temp 1697062800 26.0*  *# Lấy dữ liệu từ đầu đến thời điểm 1697062800*  *TS.RANGE sensor:temp 0 1697062800*  *# Thêm giá trị mới tại thời điểm hiện tại (09:33 AM +07, 14/09/2025, Unix: 1726287180)*  *TS.ADD sensor:temp 1726287180 26.5*  *# Xóa dữ liệu từ 1697059200 đến 1697062800*  *TS.DEL sensor:temp 1697059200 1697062800*  *# Thêm giá trị nhiệt độ khác tại thời điểm hiện tại*  *TS.ADD sensor:temp 1726287200 26.7*  *# Lấy dữ liệu trong khoảng 24 giờ gần nhất*  *TS.RANGE sensor:temp -1726200780 1726287180* |

#### Ứng dụng thực tế

* Theo dõi IoT: Lưu trữ dữ liệu nhiệt độ hoặc độ ẩm từ cảm biến.
* Phân tích tài chính: Ghi nhận giá cổ phiếu theo thời gian.
* Giám sát server: Theo dõi CPU hoặc bộ nhớ theo thời gian thực.
* Dự báo thời tiết: Lưu trữ và phân tích dữ liệu khí hậu.

## Truy vấn Full-Text Search (FT) trong Redis

### Giới thiệu

Truy vấn Full-Text Search (FT) trong Redis, được cung cấp qua module RediSearch, cho phép thực hiện tìm kiếm văn bản toàn diện trên dữ liệu được lưu trữ dưới dạng hash, JSON hoặc text.

Tương tự như aggregation trong MongoDB (sử dụng $match, $project, $sort để lọc, biến đổi và sắp xếp dữ liệu), FT kết hợp tìm kiếm văn bản với các bộ lọc, sắp xếp và phân trang để xử lý dữ liệu lớn một cách hiệu quả. Nó hỗ trợ tìm kiếm fuzzy, prefix, wildcard, và các toán tử logic, giúp Redis trở thành công cụ mạnh mẽ cho tìm kiếm thời gian thực mà không cần engine bên ngoài như Elasticsearch. Để sử dụng FT, cần cài đặt module RediSearch trên Redis server.

### Lệnh cơ bản

Để sử dụng FT, trước tiên cần tạo index bằng lệnh FT.CREATE để định nghĩa schema cho dữ liệu. Sau đó, sử dụng FT.SEARCH để thực hiện truy vấn, tương tự như aggregation pipeline trong MongoDB với các stage lọc và sắp xếp.

|  |  |  |
| --- | --- | --- |
| Chỉ thị | Cú pháp / Giá trị mặc định | Mô tả |
| FT.CREATE | FT.CREATE index\_name ON {HASH|JSON|TAG} PREFIX {n} prefix SCHEMA {field type [options]} | Tạo index cho tìm kiếm, định nghĩa schema cho các trường (TEXT cho văn bản, NUMERIC cho số, TAG cho phân loại). |
| FT.SEARCH | FT.SEARCH index\_name query [options] | Thực hiện tìm kiếm văn bản với bộ lọc, sắp xếp và phân trang (tương tự $match + $sort + $limit trong MongoDB). |
| FT.EXPLAIN | FT.EXPLAIN index\_name query | Giải thích kế hoạch thực thi truy vấn, hiển thị cách Redis xử lý (tương tự explain trong MongoDB). |
| FT.DROPINDEX | FT.DROPINDEX index\_name [DD] | Xóa index (DD để xóa dữ liệu gốc). |

### Ứng dụng thực tế

* Tìm kiếm văn bản trong catalog sản phẩm: Tạo index cho trường title và description, sử dụng FT.SEARCH với query “@title:smartphone @price:[100 500]” để lọc sản phẩm theo từ khóa và giá, tương tự aggregation pipeline trong MongoDB với $match trên text và range.
* Phân tích log hoặc bài viết: Sử dụng FT.SEARCH với fuzzy search (“hello~”) để tìm kiếm gần đúng, kết hợp SORTBY để sắp xếp theo ngày, giống như $sort và $regex trong MongoDB aggregation.
* Tìm kiếm theo vị trí địa lý: Kết hợp với trường GEO trong index để query “@loc:[lon lat radius km]”, tương tự geospatial query trong MongoDB với $geoWithin.
* Tìm kiếm thời gian thực với phân trang: Sử dụng LIMIT và OFFSET trong FT.SEARCH để phân trang kết quả, kết hợp RETURN để chọn trường trả về, giống như $skip, $limit và $project trong aggregation.

### Cách sử dụng FT

Để sử dụng FT trong Redis một cách hiệu quả, làm theo các bước sau:

1. Cài đặt RediSearch:
   * Cài module RediSearch từ nguồn chính thức hoặc thông qua package manager như *redis-stack-server*.
   * Khởi động Redis với module: *redis-server --loadmodule* */path/to/redisearch.so*.
2. Tạo Index:
   * Xác định dữ liệu cần index (hash, JSON, hoặc tag) và tạo index với *FT.CREATE*. Ví dụ, index cho sách với các trường title (TEXT) và published\_at (NUMERIC).
   * Đảm bảo thêm prefix phù hợp để tránh xung đột key.
3. Thêm dữ liệu:
   * Sử dụng lệnh *HSET* (cho hash) hoặc *JSON.SET* (cho JSON) để thêm dữ liệu vào key có prefix khớp với index.
   * Dữ liệu sẽ tự động được index hóa khi thêm.
4. Thực hiện truy vấn:
   * Sử dụng FT.SEARCH với cú pháp query:
     + Tìm kiếm cơ bản: “*term*” (tìm từ “*term*” trong tất cả trường TEXT).
     + Bộ lọc:
       1. *@field:[min max]* (lọc số hoặc khoảng thời gian)
       2. *@field:{value}* (lọc tag).
     + Toán tử: *+term1 -term2 (AND, NOT), term1 term2 (OR)*.
     + Sắp xếp: *SORTBY field [ASC\|DESC]*.
     + Phân trang: *LIMIT offset count*.
   * Ví dụ: *FT.SEARCH idx:books “@title:lord @published\_at:[1900 2000]” SORTBY published\_at DESC LIMIT 0 5*.
5. Tối ưu hóa hiệu suất:
   * Chỉ index các trường cần thiết để giảm tải bộ nhớ.
   * Sử dụng *SORTABLE* cho các trường cần sắp xếp để tăng tốc độ query.
   * Kiểm tra kế hoạch thực thi với *FT.EXPLAIN* để tối ưu query.
6. Quản lý Index:
   * Cập nhật schema hoặc xóa index với *FT.DROPINDEX* khi cần (thêm [*DD*] để xóa dữ liệu gốc).
   * Theo dõi kích thước index qua *FT.INFO index\_name*.

### Ví dụ minh họa

|  |
| --- |
| # Tạo index cho dữ liệu sách (tương tự định nghĩa schema trong MongoDB)  FT.CREATE idx:books ON HASH PREFIX 1 book: SCHEMA title TEXT SORTABLE body TEXT published\_at NUMERIC SORTABLE  # Thêm dữ liệu sách (tương tự insert documents)  HSET book:1 title “Harry Potter and the Sorcerer's Stone” body “A wizard boy discovers his powers” published\_at 1997  HSET book:2 title “The Fellowship of the Ring” body “A fellowship embarks on a quest to destroy a ring” published\_at 1954  HSET book:3 title “1984” body “A dystopian novel about surveillance and totalitarianism” published\_at 1949  # Tìm kiếm sách chứa “wizard” (tương tự $match với text search)  FT.SEARCH idx:books “wizard” RETURN 3 title body published\_at  # Tìm kiếm với bộ lọc năm xuất bản > 1950 và sắp xếp theo năm giảm dần (tương tự $match + $sort + $limit)  FT.SEARCH idx:books “@published\_at:[1950 +inf] wizard” SORTBY published\_at DESC RETURN 2 title published\_at LIMIT 0 2  # Giải thích truy vấn (tương tự explain plan)  FT.EXPLAIN idx:books “@published\_at:[1950 +inf] wizard”  # Xóa index nếu cần (tương tự drop collection)  FT.DROPINDEX idx:books |

## Một số thao tác quản trị cơ bản

### Kiểm tra thông tin server

#### Giới thiệu

Redis cung cấp các lệnh để kiểm tra trạng thái và hiệu suất server, giúp quản trị viên giám sát hệ thống.

#### Lệnh cơ bản

|  |  |  |  |
| --- | --- | --- | --- |
| Lệnh | Cú pháp | Mô tả | Ví dụ |
| INFO | INFO [section] | Hiển thị thông tin chi tiết về server (CPU, RAM, kết nối, …) | INFO SERVER |
| MONITOR | MONITOR | Theo dõi toàn bộ lệnh client gửi đến Redis theo thời gian thực | MONITOR |

#### Ứng dụng thực tế

* Giám sát hiệu suất hệ thống: Kiểm tra mức sử dụng bộ nhớ, CPU, số lượng kết nối, và thời gian hoạt động để tối ưu hóa tài nguyên. Ví dụ, một quản trị viên có thể dùng INFO để phát hiện tình trạng quá tải kết nối.
* Gỡ lỗi ứng dụng: Sử dụng MONITOR để theo dõi các lệnh được gửi từ ứng dụng, giúp phát hiện các truy vấn chậm hoặc sai logic.
* Phân tích hiệu suất: Dùng INFO để thu thập số liệu về số lượng lệnh được xử lý mỗi giây, từ đó đánh giá khả năng mở rộng của hệ thống.
* Kiểm tra trạng thái server: Đảm bảo server Redis đang hoạt động bình thường trước khi triển khai bản cập nhật hoặc bảo trì.

### Quản lý cơ sở dữ liệu

#### Giới thiệu

Redis hỗ trợ nhiều cơ sở dữ liệu logic (mặc định là 16, từ 0 đến 15), cho phép quản lý dữ liệu riêng biệt.

#### Lệnh cơ bản

|  |  |  |  |
| --- | --- | --- | --- |
| Lệnh | Cú pháp | Mô tả | Ví dụ |
| SELECT | SELECT index | Chuyển sang cơ sở dữ liệu logic theo index (0–15 mặc định) | SELECT 1 |
| DBSIZE | DBSIZE | Tra cứu số lượng key trong cơ sở dữ liệu. | DBSIZE |

#### Ứng dụng thực tế

* Phân tách dữ liệu theo ứng dụng: Sử dụng các cơ sở dữ liệu khác nhau cho các ứng dụng, như cơ sở dữ liệu 0 cho web, cơ sở dữ liệu 1 cho ứng dụng di động.
* Quản lý tài nguyên server: Dùng DBSIZE để kiểm tra số lượng key, từ đó đánh giá mức độ sử dụng và lên kế hoạch mở rộng bộ nhớ.
* Tối ưu hóa hiệu suất: Tách biệt dữ liệu của các khách hàng hoặc dự án để giảm xung đột truy cập và tăng tốc độ truy xuất.
* Kiểm tra trước khi sao lưu: Dùng DBSIZE để xác định khối lượng dữ liệu cần sao lưu hoặc di chuyển sang server khác.

### Xóa dữ liệu

#### Giới thiệu

Redis cung cấp các lệnh để xóa dữ liệu một cách linh hoạt, đảm bảo quản lý hiệu quả tài nguyên.

#### Lệnh cơ bản

|  |  |  |  |
| --- | --- | --- | --- |
| Lệnh | Cú pháp | Mô tả | Ví dụ |
| DEL | DEL key [key ...] | Xóa một hoặc nhiều key ngay lập tức (synchronous) | DEL mykey mylist |
| UNLINK | UNLINK key [key ...] | Xóa key theo cách không đồng bộ (asynchronous), tránh block server khi key lớn | UNLINK myhash |

#### Ứng dụng thực tế

* Dọn dẹp dữ liệu hết hạn: Xóa các key phiên người dùng đã hết hạn hoặc dữ liệu cache không còn cần thiết, như giỏ hàng tạm thời.
* Tối ưu hiệu suất server: Sử dụng UNLINK để xóa các key lớn (như danh sách hoặc hash lớn) mà không làm gián đoạn hoạt động server.
* Quản lý bộ nhớ: Xóa định kỳ các key không còn sử dụng để giải phóng bộ nhớ, đặc biệt trong các hệ thống có lưu lượng truy cập cao.
* Bảo trì hệ thống: Dùng DEL hoặc UNLINK để xóa dữ liệu thử nghiệm hoặc dữ liệu lỗi trong quá trình phát triển ứng dụng.

### Kiểm tra key và bộ nhớ

Một số câu lệnh dùng để kiểm tra key và bộ nhớ:

|  |  |  |  |
| --- | --- | --- | --- |
| Lệnh | Cú pháp | Mô tả | Ví dụ |
| TYPE | TYPE key | Xác định kiểu dữ liệu của key | TYPE mystring → string |
| OBJECT encoding | OBJECT encoding key | Xem cách Redis lưu trữ dữ liệu bên trong (embstr, int, ziplist, …) | OBJECT encoding mystring |
| OBJECT idletime | OBJECT idletime key | Thời gian (giây) key chưa được truy cập | OBJECT idletime mystring |
| MEMORY USAGE | MEMORY USAGE key | Dung lượng bộ nhớ (byte) mà key đang chiếm | MEMORY USAGE mystring → 50 |

## Kết luận

Chương 1 đã trình bày toàn cảnh về Redis, từ những khái niệm nền tảng đến kiến trúc, đặc điểm và cách triển khai thực tế. Redis nổi bật với cơ chế lưu trữ dữ liệu trong bộ nhớ, mang lại tốc độ xử lý vượt trội và hỗ trợ nhiều kiểu dữ liệu linh hoạt. Các tính năng quan trọng như Persistence, Replication, Cluster và Sentinel không chỉ giúp hệ thống đảm bảo độ tin cậy và khả năng mở rộng mà còn mở ra nhiều ứng dụng thiết thực trong quản lý phiên, bộ nhớ đệm hay xử lý hàng đợi.

Song song đó, phần nội dung về cài đặt và sử dụng các công cụ hỗ trợ đã giúp việc quản trị Redis trở nên trực quan, dễ dàng và hiệu quả hơn. Những thao tác cơ bản như kiểm tra thông tin hệ thống, quản lý cơ sở dữ liệu hay giám sát tài nguyên cũng được làm rõ, góp phần xây dựng nền tảng vững chắc cho người đọc trước khi đi sâu vào các chương sau. Như vậy, chương 1 không chỉ cung cấp cái nhìn toàn diện về Redis mà còn đặt nền móng quan trọng để nghiên cứu các giải pháp ứng dụng và tối ưu hiệu suất trong thực tiễn.

# PHÂN TÍCH VÀ THIẾT KẾ CƠ SỞ DỮ LIỆU NoSQL

## Bối cảnh

Trong bối cảnh chuyển đổi số diễn ra mạnh mẽ, việc quản lý và khai thác cơ hội việc làm đã trở thành yêu cầu không thể thiếu đối với cả người lao động và doanh nghiệp. Hệ thống quản lý tuyển dụng hiện đại đóng vai trò then chốt như một cầu nối hiệu quả, giúp tối ưu hóa toàn bộ quy trình từ đăng tải tin tuyển dụng, tiếp nhận hồ sơ ứng tuyển cho đến các giai đoạn sàng lọc và phỏng vấn. Theo số liệu thống kê từ Bộ Lao động - Thương binh và Xã hội Việt Nam, năm 2024 đã ghi nhận hơn 1,5 triệu lao động trẻ gia nhập thị trường việc làm. Con số ấn tượng này đặt ra thách thức lớn cho các nền tảng tuyển dụng trong việc xử lý khối lượng dữ liệu khổng lồ, đồng thời đảm bảo tốc độ phản hồi và độ tin cậy ở mức cao nhất.

## Phân tích thị trường hiện tại

Khảo sát các nền tảng tuyển dụng phổ biến tại Việt Nam cho thấy mỗi hệ thống đều có những điểm mạnh và hạn chế riêng. TopCV nổi bật với khả năng matching tự động thông minh, tuy nhiên vẫn chưa tối ưu về xử lý dữ liệu real-time. CareerLink được đánh giá cao về giao diện thân thiện với doanh nghiệp vừa và nhỏ, nhưng hiệu suất xử lý giảm đáng kể khi quy mô dữ liệu tăng cao. VietnamWorks sở hữu lượng dữ liệu lớn nhưng đi kèm với chi phí cao và khả năng tùy chỉnh hạn chế.

## Giải pháp đề xuất

Để khắc phục những hạn chế trên, hệ thống đề xuất tận dụng sức mạnh của MongoDB - cơ sở dữ liệu NoSQL hướng tài liệu (document-oriented), cho phép lưu trữ dữ liệu phi cấu trúc một cách linh hoạt với khả năng mở rộng theo chiều ngang. Kết hợp với Redis - hệ thống caching in-memory hiệu suất cao, hỗ trợ xử lý real-time và quản lý session. Kiến trúc này tập trung vào hai collection chính với thiết kế nhúng (embedded) tối ưu: User (nhúng thông tin Company) và Job (nhúng danh sách Applicant).

## Phân tích yêu cầu hệ thống

### Yêu cầu chức năng

Hệ thống quản lý tuyển dụng được thiết kế để phục vụ ba nhóm đối tượng chính với các chức năng đặc thù:

**Ứng viên (Candidate):** Đây là vai trò mặc định khi người dùng đăng ký tài khoản mới. Ứng viên có khả năng tìm kiếm việc làm theo nhiều tiêu chí đa dạng như vị trí, địa điểm, mức lương và kinh nghiệm yêu cầu. Họ có thể nộp hồ sơ ứng tuyển trực tuyến kèm thư xin việc và tài liệu đính kèm, theo dõi trạng thái xét duyệt theo thời gian thực qua các giai đoạn sàng lọc và phỏng vấn, cũng như cập nhật hồ sơ cá nhân bao gồm kỹ năng, kinh nghiệm làm việc và trình độ học vấn.

**Nhà tuyển dụng (Recruiter):** Người dùng có thể chuyển đổi từ vai trò Candidate sang Recruiter thông qua quy trình xác minh của quản trị viên. Khi gửi yêu cầu chuyển đổi, người dùng phải cung cấp đầy đủ thông tin doanh nghiệp bao gồm tên công ty, website chính thức, email doanh nghiệp, số điện thoại liên hệ, quy mô nhân sự, lĩnh vực kinh doanh, địa chỉ trụ sở và giới thiệu về công ty. Nếu công ty đã tồn tại trong hệ thống, người dùng chỉ cần liên kết với công ty hiện có. Sau khi được phê duyệt, Recruiter có quyền đăng tải tin tuyển dụng, quản lý danh sách ứng viên, thực hiện sàng lọc hồ sơ thông qua hệ thống chấm điểm, lập lịch phỏng vấn và xuất các báo cáo thống kê về hiệu quả tuyển dụng.

**Quản trị viên (Admin):** Có toàn quyền quản lý người dùng trên hệ thống, phê duyệt các yêu cầu chuyển đổi vai trò từ Candidate sang Recruiter, xác minh thông tin doanh nghiệp, giám sát hoạt động tổng thể của nền tảng và tạo các báo cáo tổng hợp phục vụ công tác quản lý.

### Yêu cầu phi chức năng

**Hiệu năng:** Hệ thống cần đảm bảo thời gian phản hồi tối đa 200ms cho các truy vấn thông thường và hỗ trợ đồng thời tối thiểu 1.000 người dùng mà không giảm chất lượng dịch vụ. Đây là yêu cầu quan trọng trong bối cảnh thị trường tuyển dụng với lưu lượng truy cập cao vào giờ cao điểm.

**Bảo mật:** Triển khai mã hóa bcrypt cho mật khẩu người dùng với cost factor tối thiểu là 10, áp dụng xác thực JWT (JSON Web Token) với thời gian sống hợp lý cho mỗi phiên làm việc, mã hóa HTTPS cho mọi giao tiếp giữa client và server, và tuân thủ Nghị định 13/2023/NĐ-CP về bảo vệ dữ liệu cá nhân tại Việt Nam.

**Khả năng mở rộng:** Hệ thống phải có khả năng mở rộng theo chiều ngang (horizontal scaling) khi số lượng người dùng và dữ liệu tăng trưởng. MongoDB hỗ trợ sharding tự động để phân tán dữ liệu và replica set để đảm bảo tính sẵn sàng cao.

**Khả dụng:** Đảm bảo uptime 99.9% trong năm, tương đương tối đa 8.76 giờ downtime, với phương án backup tự động hàng ngày và kế hoạch disaster recovery cho phép khôi phục dữ liệu trong vòng 4 giờ (RTO - Recovery Time Objective).

## Lựa chọn công nghệ: So sánh SQL và NoSQL

### Phân tích đặc thù của hệ thống tuyển dụng

Trước khi đưa ra quyết định về công nghệ cơ sở dữ liệu, cần phân tích đặc điểm của hệ thống tuyển dụng:

**Tính đa dạng của dữ liệu:** Thông tin ứng viên và yêu cầu công việc rất đa dạng và thay đổi liên tục. Một vị trí trong ngành công nghệ có thể yêu cầu danh sách kỹ năng kỹ thuật chi tiết, trong khi vị trí marketing lại tập trung vào portfolio sáng tạo. Mỗi ngành nghề có những yêu cầu riêng biệt, khó có thể chuẩn hóa thành schema cố định.

**Mẫu truy vấn chủ đạo:** Hơn 80% các thao tác trong hệ thống là đọc dữ liệu - tìm kiếm công việc, xem hồ sơ ứng viên, hiển thị thông tin công ty. Các thao tác ghi (tạo hồ sơ, cập nhật trạng thái) chiếm tỷ lệ nhỏ hơn. Điều này ảnh hưởng đến việc lựa chọn mô hình tối ưu cho read-heavy workload.

**Nhu cầu mở rộng:** Khi hệ thống phát triển, số lượng người dùng và tin tuyển dụng có thể tăng trưởng nhanh chóng. Một nền tảng tuyển dụng thành công có thể có hàng triệu người dùng và hàng chục triệu tin tuyển dụng, đòi hỏi khả năng scale hiệu quả.

### So sánh SQL và NoSQL trong bối cảnh tuyển dụng

**Tính linh hoạt của schema:**

SQL với schema cứng nhắc yêu cầu định nghĩa trước cấu trúc bảng. Khi cần thêm trường mới (ví dụ: thêm trường “remote\_salary” cho các vị trí làm việc từ xa), cần thực hiện migration ALTER TABLE trên toàn bộ dữ liệu, gây downtime hoặc lock table. Đối với dữ liệu đa dạng như hồ sơ ứng viên, việc sử dụng nhiều bảng phụ (candidate\_skills, candidate\_certifications, candidate\_education) hoặc các trường nullable dẫn đến sự phức tạp không cần thiết.

MongoDB cho phép mỗi document có cấu trúc riêng biệt. Một ứng viên ngành IT có thể có mảng certifications chứa các chứng chỉ kỹ thuật, trong khi ứng viên ngành nghệ thuật có thể có trường portfolio chứa link các tác phẩm. Không cần migration khi thêm trường mới, chỉ cần cập nhật document cụ thể.

**Hiệu suất truy vấn:**

Trong SQL, để hiển thị một tin tuyển dụng đầy đủ cần JOIN nhiều bảng:

* JOIN với bảng Company để lấy thông tin công ty
* JOIN với bảng Applications để lấy danh sách ứng viên
* JOIN với bảng Users để lấy thông tin chi tiết mỗi ứng viên
* JOIN với bảng Job\_Skills để lấy yêu cầu kỹ năng

Mỗi JOIN tốn chi phí xử lý, đặc biệt khi dữ liệu lớn. Index có thể cải thiện nhưng không loại bỏ hoàn toàn overhead.

MongoDB với embedded documents có thể trả về toàn bộ thông tin trong một lần truy vấn duy nhất. Job document đã chứa sẵn mảng applicants với thông tin snapshot của mỗi ứng viên, không cần JOIN. Điều này đặc biệt quan trọng khi recruiter xem danh sách ứng viên - thao tác được thực hiện liên tục và cần tốc độ cao.

**Khả năng mở rộng:**

SQL truyền thống mở rộng theo chiều dọc (vertical scaling) - nâng cấp CPU, RAM, disk của server. Khi đạt giới hạn phần cứng, việc scale tiếp trở nên rất tốn kém. Sharding trong SQL phức tạp, thường cần công cụ bên thứ ba và yêu cầu thay đổi code ứng dụng đáng kể.

MongoDB hỗ trợ sharding tự động (horizontal scaling) - phân tán dữ liệu trên nhiều server theo shard key. Khi cần tăng capacity, chỉ cần thêm shard mới vào cluster mà không cần thay đổi code. MongoDB tự động cân bằng dữ liệu giữa các shard.

### Hạn chế của NoSQL và cách khắc phục

**Tính nhất quán dữ liệu:**

MongoDB ban đầu không đảm bảo ACID ở cấp độ multi-document như SQL. Tuy nhiên, từ phiên bản 4.0, MongoDB đã hỗ trợ multi-document transactions, cho phép thực hiện các thao tác phức tạp cần tính nhất quán. Trong hệ thống tuyển dụng, transactions được áp dụng khi cần cập nhật đồng thời nhiều documents quan trọng, ví dụ khi thay đổi trạng thái ứng tuyển và gửi thông báo.

**Dữ liệu trùng lặp:**

Với embedded documents, thông tin công ty được nhúng trong User document của mỗi recruiter có thể dẫn đến dư thừa. Giải pháp là chỉ lưu tham chiếu (companyId) trong User, và cache thông tin công ty quan trọng trong Redis để tăng tốc độ truy xuất. Khi thông tin công ty thay đổi, chỉ cần cập nhật một bản ghi chính và invalidate cache.

**Truy vấn phân tích phức tạp:**

Các báo cáo thống kê đa chiều (ví dụ: phân tích xu hướng tuyển dụng theo thời gian, ngành nghề, địa lý) khó thực hiện hơn so với SQL. Giải pháp là sử dụng MongoDB Aggregation Pipeline cho các báo cáo thời gian thực, và đồng bộ dữ liệu định kỳ sang data warehouse (PostgreSQL hoặc BigQuery) cho các phân tích sâu offline.

## Thiết kế cơ sở dữ liệu NoSQL

### Triết lý thiết kế: Embedded vs Reference

Trong MongoDB, có hai cách chính để mô hình hóa mối quan hệ giữa các thực thể:

**Embedding (Nhúng):** Dữ liệu con được lưu trực tiếp bên trong document cha dưới dạng nested document hoặc array. Phù hợp khi:

* Dữ liệu con có mối quan hệ chặt chẽ với dữ liệu cha (relationship “1-to-few”)
* Dữ liệu thường được truy vấn cùng nhau
* Dữ liệu con có kích thước giới hạn (MongoDB giới hạn document size là 16MB)

Trong hệ thống, danh sách applicants được nhúng trực tiếp vào Job document vì khi xem một tin tuyển dụng, recruiter thường muốn xem ngay danh sách ứng viên và trạng thái của họ.

**Referencing (Tham chiếu):** Lưu ObjectId của document khác để tham chiếu, giống foreign key trong SQL. Phù hợp khi:

* Dữ liệu có mối quan hệ lỏng lẻo
* Dữ liệu được truy cập độc lập
* Dữ liệu con có thể phát triển không giới hạn (relationship “1-to-many” hoặc “many-to-many”)

Chúng tôi lưu companyId như một reference trong User document thay vì nhúng toàn bộ company object để tránh dư thừa khi nhiều recruiter thuộc cùng một công ty.

### Mô hình dữ liệu chi tiết

Hệ thống sử dụng hai collections chính:

* **Collection USER**

Collection này lưu trữ tất cả người dùng của hệ thống, bao gồm candidate, recruiter và admin. Các trường chính:

* \_id: ObjectId tự động sinh bởi MongoDB, đóng vai trò Primary Key
* email: String duy nhất, sử dụng cho đăng nhập và liên hệ
* passwordHash: Mật khẩu đã được mã hóa bằng bcrypt
* role: Enum ['candidate', 'recruiter', 'admin'], mặc định là 'candidate'
* profile: Object chứa thông tin cá nhân dành cho candidate, bao gồm các mảng nested cho education, experience, skills, certifications
* company: Object chứa thông tin công ty (chỉ có khi role='recruiter')
* roleChangeRequest: Object theo dõi yêu cầu chuyển đổi vai trò

Thiết kế này cho phép một user chuyển đổi linh hoạt giữa các vai trò mà không cần migrate sang bảng khác. Trường profile với cấu trúc lồng nhau phù hợp với đặc điểm thông tin ứng viên - những dữ liệu này thường được truy vấn cùng nhau khi xem hồ sơ.

* **Collection JOB**

Collection này lưu trữ tin tuyển dụng và danh sách ứng viên. Các trường chính:

* \_id: ObjectId của tin tuyển dụng
* companyId, recruiterId: Reference đến User collection
* Thông tin công việc: title, description, salary, requirements, responsibilities, benefits
* Yêu cầu: experience, education, employmentType, level
* skills: Mảng các kỹ năng yêu cầu với trọng số (weight) cho thuật toán matching
* applicants: Mảng embedded documents chứa thông tin chi tiết của từng ứng viên

Mỗi phần tử trong mảng applicants là một subdocument đầy đủ với:

* applicantSnapshot: Bản chụp thông tin ứng viên tại thời điểm apply (quan trọng vì ứng viên có thể cập nhật profile sau đó)
* screening: Kết quả sàng lọc tự động với điểm số matching
* interviews: Mảng các vòng phỏng vấn với feedback chi tiết
* timeline: Lịch sử thay đổi trạng thái

Thiết kế này tối ưu cho trường hợp recruiter xem chi tiết một tin tuyển dụng cùng danh sách ứng viên - chỉ cần một query duy nhất thay vì phải JOIN nhiều bảng.

* **Indexes và tối ưu hóa**

**Indexes trên User collection:**

* email: Unique index để đảm bảo không trùng lặp và tăng tốc login
* role: Non-unique index để lọc user theo vai trò
* company.\_id: Non-unique index để tìm tất cả recruiter của một công ty

**Indexes trên Job collection:**

* companyId, recruiterId: Non-unique indexes cho lookup
* status: Non-unique index để lọc job đang active
* categories, skills.name: Array indexes để hỗ trợ tìm kiếm theo ngành nghề và kỹ năng
* locations.city: Non-unique index cho tìm kiếm theo địa điểm
* Compound index (startDate, endDate) để tối ưu query tìm job trong khoảng thời gian

**Denormalization cho hiệu năng:**

Trường stats trong Job document được denormalize (tính toán trước và lưu lại) thay vì tính on-the-fly:

|  |
| --- |
| stats: {  totalApplicants: 15,  pendingCount: 5,  interviewingCount: 4,  avgScreeningScore: 76.5  } |

Mỗi khi có thay đổi trạng thái ứng viên, các counters này được cập nhật bằng atomic operations:

|  |
| --- |
| db.jobs.updateOne(  { \_id: jobId },  {  $inc: {  “stats.pendingCount”: -1,  “stats.interviewingCount”: 1  }  }  ) |

## So sánh hiệu năng truy vấn

### Kịch bản 1: Hiển thị chi tiết tin tuyển dụng với danh sách ứng viên

**Với SQL (thiết kế normalized):**

|  |
| --- |
| *-- Query 1: Lấy thông tin job*  SELECT \* FROM jobs WHERE id = ?;  *-- Query 2: Lấy thông tin công ty*  SELECT \* FROM companies WHERE id = ?;  *-- Query 3: Lấy danh sách applications*  SELECT \* FROM applications WHERE job\_id = ?;  *-- Query 4: Với mỗi application, lấy thông tin candidate (giả sử 15 ứng viên)*  SELECT \* FROM users WHERE id IN (?, ?, ..., ?);  *-- Query 5: Lấy skills của mỗi candidate*  SELECT \* FROM user\_skills WHERE user\_id IN (?, ?, ..., ?); |

Tổng: **5 queries với nhiều JOINs**, mỗi query phải scan indexes và thực hiện JOIN. Với 15 ứng viên, query 5 phải xử lý khoảng 75-150 records (trung bình mỗi ứng viên 5-10 skills).

Thời gian ước tính: **150-300ms** (tùy thuộc vào network latency và database load)

**Với MongoDB (thiết kế embedded):**

|  |
| --- |
| db.jobs.findOne({ \_id: jobId }) |

Chỉ **1 query duy nhất** trả về toàn bộ Job document bao gồm mảng applicants với đầy đủ thông tin. MongoDB đọc một document liên tục từ disk, không cần JOIN.

Thời gian ước tính: **10-30ms**

**Kết luận:** MongoDB nhanh hơn **5-10 lần** cho trường hợp này.

### Kịch bản 2: Tìm kiếm công việc theo nhiều tiêu chí

**Yêu cầu:** Tìm job trong ngành IT, ở TP.HCM, yêu cầu skill “React”, mức lương 20-40 triệu, đang active.

**Với SQL:**

|  |
| --- |
| SELECT j.\*, c.name as company\_name, c.logo  FROM jobs j  JOIN companies c ON j.company\_id = c.id  JOIN job\_skills js ON j.id = js.job\_id  WHERE j.status = 'published'  AND j.salary\_min >= 20000000  AND j.salary\_max <= 40000000  AND j.location\_city = 'Hồ Chí Minh'  AND js.skill\_name = 'React'  AND j.category = 'IT'  ORDER BY j.created\_at DESC  LIMIT 20; |

JOIN giữa jobs, companies và job\_skills. Với index hợp lý, query vẫn phải scan nhiều records.

Thời gian ước tính: **50-100ms**

**Với MongoDB:**

|  |
| --- |
| db.jobs.find({  status: 'published',  'salary.min': { $gte: 20000000 },  'salary.max': { $lte: 40000000 },  'locations.city': 'Hồ Chí Minh',  'skills.name': 'React',  categories: 'IT'  })  .sort({ createdAt: -1 })  .limit(20) |

Không cần JOIN vì thông tin công ty có thể được cache. Với compound index phù hợp, MongoDB scan rất hiệu quả.

Thời gian ước tính: **20-40ms**

**Kết luận:** MongoDB nhanh hơn **2-3 lần** nhờ tránh JOIN và tận dụng document model.

### Kịch bản 3: Cập nhật trạng thái ứng viên

**Với SQL:**

|  |
| --- |
| BEGIN TRANSACTION;  *-- Update application status*  UPDATE applications  SET status = 'interviewing',  updated\_at = NOW()  WHERE id = ?;  *-- Insert timeline record*  INSERT INTO application\_timeline (application\_id, status, changed\_at, changed\_by)  VALUES (?, 'interviewing', NOW(), ?);  *-- Update job stats*  UPDATE jobs  SET interviewing\_count = interviewing\_count + 1,  pending\_count = pending\_count - 1  WHERE id = ?;  COMMIT; |

Cần **3 queries trong transaction** để đảm bảo tính nhất quán.

Thời gian ước tính: **30-60ms**

**Với MongoDB:**

|  |
| --- |
| db.jobs.updateOne(  {  \_id: jobId,  'applicants.\_id': applicationId  },  {  $set: {  'applicants.$.status': 'interviewing',  'applicants.$.lastStatusUpdate': new Date()  },  $push: {  'applicants.$.timeline': {  status: 'interviewing',  changedAt: new Date(),  changedBy: userId  }  },  $inc: {  'stats.interviewingCount': 1,  'stats.pendingCount': -1  }  }  ) |

Chỉ **1 atomic operation** cập nhật nhiều trường trong cùng một document. MongoDB đảm bảo atomicity ở document level.

Thời gian ước tính: **10-20ms**

**Kết luận:** MongoDB nhanh hơn **2-3 lần** nhờ atomic operations và tránh transaction overhead.

### Tổng kết so sánh hiệu năng

|  |  |  |  |
| --- | --- | --- | --- |
| Kịch bản | SQL | MongoDB | Tỷ lệ |
| Hiển thị job + applicants | 150-300ms | 10-30ms | **5-10x** |
| Tìm kiếm job | 50-100ms | 20-40ms | **2-3x** |
| Cập nhật trạng thái | 30-60ms | 10-20ms | **2-3x** |

MongoDB có lợi thế rõ rệt trong các trường hợp:

* Truy vấn dữ liệu có mối quan hệ 1-to-many cần hiển thị cùng nhau
* Read-heavy workload với ít write
* Dữ liệu có cấu trúc linh hoạt, thay đổi thường xuyên

SQL vẫn phù hợp hơn cho:

* Các truy vấn phân tích phức tạp với nhiều aggregations
* Dữ liệu có mối quan hệ many-to-many phức tạp
* Yêu cầu ACID transactions nghiêm ngặt across nhiều entities

## Chiến lược caching với Redis

Để tối ưu thêm hiệu năng, hệ thống tích hợp Redis cho các mục đích:

**Session storage:** Lưu JWT tokens và user sessions với TTL (Time To Live) tự động expire.

**Query result caching:** Cache kết quả các truy vấn phổ biến như danh sách job mới nhất, top companies, với invalidation strategy khi có cập nhật.

**Rate limiting:** Giới hạn số lượng requests từ mỗi user/IP để tránh abuse.

**Real-time counters:** Đếm lượt xem job, số ứng viên mới trong Redis trước khi sync về MongoDB định kỳ, giảm write load lên database chính.

**Leaderboard:** Rankings như “Top 10 jobs có nhiều applicants nhất” sử dụng Redis Sorted Sets để cập nhật và query cực nhanh.

Với kiến trúc MongoDB + Redis, hệ thống đạt được sự cân bằng tối ưu giữa hiệu năng, tính linh hoạt và khả năng mở rộng, đáp ứng tốt các yêu cầu của một nền tảng tuyển dụng hiện đại.

## Kết luận

Chương 2 đã trình bày chi tiết quá trình phân tích và thiết kế cơ sở dữ liệu NoSQL cho hệ thống quản lý tuyển dụng. Việc lựa chọn MongoDB thay vì SQL truyền thống dựa trên những phân tích cụ thể về đặc thù của hệ thống: tính đa dạng của dữ liệu, mẫu truy vấn read-heavy và nhu cầu mở rộng linh hoạt.

Thiết kế hai collections chính (User và Job) với chiến lược embedding cho các mối quan hệ 1-to-many đã được chứng minh mang lại hiệu năng vượt trội, với tốc độ truy vấn nhanh hơn 2-10 lần so với mô hình SQL normalized. Kết quả so sánh cho thấy MongoDB đặc biệt phù hợp cho các trường hợp cần hiển thị dữ liệu có mối quan hệ chặt chẽ, như việc hiển thị tin tuyển dụng cùng danh sách ứng viên.

Chiến lược kết hợp MongoDB cho persistent storage và Redis cho caching tạo nên một kiến trúc cân bằng giữa hiệu năng, tính linh hoạt và khả năng mở rộng, đáp ứng tốt các yêu cầu của một nền tảng tuyển dụng hiện đại với hàng nghìn người dùng đồng thời.

# TRIỂN KHAI ỨNG DỤNG

## Tổng quan kiến trúc hệ thống

### Kiến trúc tổng thể

Hệ thống quản lý tuyển dụng được xây dựng theo mô hình kiến trúc 3 lớp (3-tier architecture), đảm bảo sự tách biệt rõ ràng giữa các tầng xử lý, giúp dễ dàng bảo trì và mở rộng trong tương lai. Kiến trúc này bao gồm ba tầng chính:

**Tầng trình diễn (Presentation Layer):** Giao diện người dùng sử dụng các framework hiện đại như React, Angular hoặc Vue.js. Tầng này chịu trách nhiệm hiển thị dữ liệu, thu thập input từ người dùng và giao tiếp với Backend thông qua RESTful APIs. Toàn bộ giao tiếp được mã hóa qua HTTPS để đảm bảo bảo mật.

**Tầng logic nghiệp vụ (Business Logic Layer):** ASP.NET Core Web API đóng vai trò trung gian, xử lý toàn bộ logic nghiệp vụ của hệ thống. Các chức năng chính bao gồm:

* Xác thực người dùng và phân quyền truy cập
* Validation dữ liệu đầu vào
* Xử lý nghiệp vụ phức tạp: matching ứng viên, tính điểm screening
* Điều phối các yêu cầu từ Frontend đến Backend services

**Tầng dữ liệu (Data Access Layer):** MongoDB đóng vai trò cơ sở dữ liệu chính để lưu trữ persistent data, trong khi Redis được sử dụng như lớp caching trung gian để tối ưu hiệu năng. Tầng này được thiết kế theo Repository Pattern, tạo ra lớp trừu tượng giữa business logic và data storage.

### Luồng hoạt động của hệ thống

Khi người dùng thực hiện một thao tác trên giao diện (ví dụ: tìm kiếm công việc), request được gửi từ Browser đến API Server thông qua HTTPS. Quá trình xử lý diễn ra theo các bước sau:

1. API Server xác thực JWT token trong request header để đảm bảo người dùng đã đăng nhập
2. Controller nhận request và chuyển đến Service Layer để xử lý business logic
3. Service Layer kiểm tra Redis Cache:

* Cache hit: Dữ liệu được trả về ngay lập tức
* Cache miss: Service gọi Repository để truy vấn MongoDB, sau đó lưu kết quả vào Redis Cache

1. Kết quả cuối cùng được serialize thành JSON và trả về cho Client

## Công nghệ và công cụ sử dụng

### Backend - ASP.NET Core Web API

Hệ thống backend được xây dựng trên nền tảng ASP.NET Core 8.0 (Long-Term Support). ASP.NET Core được lựa chọn vì những ưu điểm vượt trội: đây là framework cross-platform cho phép triển khai trên Windows, Linux và macOS, mang lại sự linh hoạt trong việc chọn môi trường production. Hiệu năng cao với Kestrel web server được Microsoft tối ưu hóa, có khả năng xử lý hàng nghìn requests đồng thời. Built-in Dependency Injection giúp quản lý lifecycle của các services một cách tự động và hiệu quả.

**Cấu trúc project theo Clean Architecture:**

|  |  |
| --- | --- |
| **Project** | **Chức năng** |
| API | Controllers, Program.cs |
| Core | Domain Models, DTOs, Interfaces (độc lập hoàn toàn) |
| Infrastructure | Repositories, Services, MongoDB/Redis integration |

### Cơ sở dữ liệu - MongoDB

MongoDB Community Server 7.0 được triển khai như primary database của hệ thống. Việc cài đặt MongoDB thực hiện thông qua installer chính thức, chọn chế độ “Complete installation” và “Install MongoDB as a Service” để tự động khởi động cùng hệ điều hành. MongoDB Compass được cài kèm theo, cung cấp giao diện trực quan để quản lý database.

**Thiết lập ban đầu:**

* Database name: RecruitmentDB
* Driver: MongoDB.Driver for .NET với LINQ provider, async operations và change streams

Cấu hình này đảm bảo ứng dụng có đủ quyền để thực hiện các thao tác CRUD và quản lý indexes nhưng không có quyền admin toàn hệ thống, tuân thủ nguyên tắc least privilege trong bảo mật.

### Caching và Session - Redis

Redis 7.2 là in-memory data store được sử dụng làm caching layer trong hệ thống. Trên Windows, Redis được triển khai thông qua Docker với lệnh docker run -d -p 6379:6379 redis:7.2-alpine, khởi động Redis server lắng nghe trên port mặc định 6379.

**Mục đích sử dụng Redis:**

|  |  |  |
| --- | --- | --- |
| Chức năng | Mô tả | TTL |
| Session Management | Lưu JWT refresh tokens | 7 ngày |
| Query Result Caching | Cache danh sách jobs, top companies | 5-10 phút |
| Rate Limiting | Giới hạn requests từ mỗi IP/user | 1 phút |
| Real-time Counters | Đếm view count, sync về MongoDB | 5 phút |

StackExchange.Redis là thư viện .NET client được tích hợp để làm việc với Redis. Trong project, RedisCacheService được implement để wrap các operations cơ bản (Get, Set, Remove, Increment) và inject vào business services thông qua Dependency Injection.

### Xác thực và bảo mật - JWT

JSON Web Token (JWT) là chuẩn mở (RFC 7519) được sử dụng cho stateless authentication trong hệ thống. Khi user đăng nhập thành công, server tạo một JWT chứa thông tin user (userId, email, role) và ký bằng secret key. Token này được trả về cho client và client lưu trữ để gửi kèm trong mỗi request tiếp theo thông qua header Authorization: Bearer <token>.

**Refresh Token Strategy:**

* Access token: Thời gian sống ngắn (60 phút) để giảm thiểu rủi ro nếu bị đánh cắp
* Refresh token: Lưu trong Redis với thời gian sống 7 ngày
* Token rotation: Mỗi lần dùng sẽ generate token mới và invalidate token cũ

**Mã hóa mật khẩu:** BCrypt.Net-Next được sử dụng với cost factor 10-12 cho production, tạo ra hash mất khoảng 100-300ms, đủ chậm để chống brute-force nhưng không ảnh hưởng trải nghiệm người dùng.

### Các công nghệ hỗ trợ

* FluentValidation: Cung cấp cách viết validation rules rõ ràng, tách biệt khỏi domain models
* Serilog: Thư viện logging mạnh mẽ với structured logging và nhiều sinks (console, file, database)
* Swashbuckle (Swagger UI): Tự động generate API documentation từ code
* MailKit: Thư viện cross-platform để gửi email xác thực, thông báo trạng thái ứng tuyển

## Cài đặt và cấu hình môi trường

### Khởi tạo project ASP.NET Core

Quá trình khởi tạo project bắt đầu với việc tạo Solution chứa nhiều projects. Sử dụng .NET CLI để tạo các projects theo Clean Architecture:

|  |
| --- |
| dotnet new sln -n RecruitmentSystem  dotnet new webapi -n API  dotnet new classlib -n Core  dotnet new classlib -n Infrastructure |

**Cấu hình appsettings.json:** File này chứa tất cả configuration settings của ứng dụng dưới dạng JSON hierarchy:

* DatabaseSettings: MongoDB connection string, database name, collection names
* RedisSettings: Connection string và instance name (prefix cho tất cả keys)
* JwtSettings: Secret key, issuer, audience, expiry times cho access và refresh token

Các sensitive information như passwords, secret keys sử dụng User Secrets trong development và environment variables trong production, không commit trực tiếp vào source control.

### Cấu hình Dependency Injection

ASP.NET Core có built-in DI container để quản lý dependencies với ba lifecycle chính:

|  |  |  |
| --- | --- | --- |
| Lifecycle | Mô tả | Sử dụng cho |
| Singleton | Instance tạo một lần duy nhất cho toàn bộ application | MongoDB client, Redis connection |
| Scoped | Instance mới cho mỗi HTTP request | Repositories, Database contexts |
| Transient | Instance mới mỗi khi service được request | Lightweight, stateless services |

Settings classes được bind từ appsettings.json và register như IOptions, cho phép inject settings vào constructors một cách type-safe.

### Tích hợp MongoDB Driver

MongoDB.Driver package cung cấp API để interact với MongoDB từ .NET. MongoClient là entry point, được khởi tạo với connection string và nên được reuse (singleton) vì nó quản lý connection pool internally.

**Repository Pattern Implementation:** Repository là abstraction layer giữa business logic và data access. Mỗi entity (User, Job, Company) có:

* Repository interface định nghĩa các operations (GetById, Create, Update, Delete, Search)
* Concrete implementation sử dụng MongoDB driver để thực hiện operations

Ưu điểm của pattern này là business logic không phụ thuộc trực tiếp vào MongoDB - nếu cần chuyển sang database khác, chỉ cần implement repository mới mà không thay đổi business logic.

### Tích hợp Redis với StackExchange.Redis

StackExchange.Redis là high-performance Redis client cho .NET. ConnectionMultiplexer quản lý connection pool đến Redis, tốn kém để khởi tạo nên được reuse (singleton).

**RedisCacheService Implementation:**

|  |  |
| --- | --- |
| Method | Chức năng |
| GetAsync<T> | Lấy data từ Redis, deserialize từ JSON → object type T |
| SetAsync<T> | Serialize object → JSON, lưu vào Redis với optional expiry |
| RemoveAsync | Xóa key khỏi Redis |
| IncrementAsync | Atomic increment counter (view counts, rate limiting) |

Service này được inject vào business services thông qua DI, cho phép dễ dàng mock trong unit tests.

### Cấu hình JWT Authentication

Middleware JWT Authentication được thêm vào pipeline để validate JWT tokens trong requests. Middleware này extract token từ Authorization header, verify signature sử dụng secret key từ configuration, validate issuer, audience và expiry time. Nếu valid, middleware populate HttpContext.User với claims từ token, cho phép các controllers truy cập user information.

**Token Generation:** Khi user login thành công, JwtSecurityTokenHandler tạo token mới với claims chứa userId, email, role. Token được ký bằng symmetric key (HMAC-SHA256), phù hợp khi issuer và consumer là cùng một service.

**Authorization Attributes:**

* [Authorize]: Require authentication
* [Authorize(Roles = “admin”)]: Restrict access chỉ cho admin role

Middleware tự động return 401 Unauthorized hoặc 403 Forbidden nếu không có quyền

## Triển khai các API Endpoints

### Authentication APIs

**POST /api/auth/register**

Endpoint này nhận email, password và basic info từ client. Quy trình xử lý diễn ra như sau:

* Validate input bằng FluentValidation: kiểm tra email format, password strength (tối thiểu 8 ký tự, có chữ hoa, số, ký tự đặc biệt)
* Check email đã tồn tại chưa bằng repository
* Hash password bằng BCrypt với cost factor 10
* Tạo User object với role mặc định “candidate” và isVerified=false
* Insert vào MongoDB và generate verification token
* Gửi email xác thực, return success response

**POST /api/auth/login**

Endpoint xử lý đăng nhập người dùng:

* Tìm user bằng email trong database
* Verify password bằng BCrypt.Verify() - so sánh password input với hash trong database
* Generate JWT access token (expiry 60 phút) và refresh token
* Lưu refresh token vào Redis với expiry 7 ngày
* Update lastLogin timestamp
* Return tokens và user info

**POST /api/auth/refresh**

Endpoint này nhận refresh token từ client, validate token tồn tại trong Redis. Nếu valid, generate access token mới và rotate refresh token - tạo token mới, invalidate token cũ trong Redis. Chiến lược này đảm bảo nếu refresh token bị đánh cắp, attacker chỉ dùng được một lần.

### User Management APIs

**GET /api/users/profile**

Endpoint yêu cầu authentication, extract userId từ JWT claims và query MongoDB để lấy User document. Dữ liệu được map sang DTO (Data Transfer Object) để hide sensitive fields như passwordHash. Response được cache trong Redis với key user:{userId}:profile và expiry 10 phút.

**PUT /api/users/profile**

Endpoint này validate input data, update User document trong MongoDB sử dụng UpdateOne() với update operators ($set cho các fields đơn giản, $push cho arrays). Sau khi update thành công, invalidate cache trong Redis để force refresh ở lần query tiếp theo, sau đó return updated profile.

### Job Management APIs

**POST /api/jobs** (Require role “recruiter”)

Quy trình tạo job mới:

* Validate job data đầu vào
* Extract companyId và recruiterId từ authenticated user
* Tạo Job document với applicants array rỗng, stats counters = 0, status “draft”
* Insert vào MongoDB, return created job với ID

**PATCH /api/jobs/{id}/publish**

Update job status từ “draft” sang “published”, set publishedAt timestamp. Invalidate các cache liên quan đến job listings để đảm bảo danh sách được cập nhật.

**GET /api/jobs/search**

Endpoint tìm kiếm job với các tính năng:

|  |  |
| --- | --- |
| **Tính năng** | **Mô tả** |
| Query parameters | keyword, location, salary range, categories |
| Filter | Build MongoDB filter bằng FilterDefinitionBuilder |
| Pagination | Skip và limit |
| Sorting | Theo createdAt descending |
| Caching | Frequently searched queries trong Redis |

### Application APIs

**POST /api/jobs/{id}/apply** (Require role “candidate”)

Endpoint xử lý ứng tuyển công việc:

* Validate jobId tồn tại và check user chưa apply job này
* Lấy snapshot thông tin candidate từ User profile
* Tạo Applicant subdocument với snapshot, coverLetter, resume URL, status “pending”
* Push vào applicants array của Job document bằng $push operator
* Increment stats.totalApplicants và stats.pendingCount
* Trigger background job để calculate screening score
* Return success response

**PATCH /api/applications/{id}/status** (Require role “recruiter”)

Endpoint cập nhật trạng thái hồ sơ ứng tuyển. Recruiter phải là người tạo job, system validate new status hợp lệ và update applicant status bằng positional operator ($) để update phần tử trong array. Timeline entry được push, relevant stats counters được update, email notification được gửi cho candidate.

### Company Management APIs

**POST /api/companies/register**

Endpoint này nhận thông tin đăng ký công ty từ client. Quy trình xử lý diễn ra như sau:

* Validate input bằng FluentValidation: kiểm tra tên công ty, địa chỉ, thông tin liên hệ
* Check tên công ty đã tồn tại chưa trong database
* Tạo Company object với các thông tin cơ bản
* Insert vào MongoDB collection companies
* Tạo một User account với role "company\_admin" cho người đại diện
* Return thông tin công ty đã đăng ký thành công

**GET /api/companies**

Endpoint lấy danh sách các công ty:

* Hỗ trợ phân trang với pageSize và pageNumber
* Cho phép filter theo tên công ty, địa điểm
* Trả về danh sách công ty với thông tin cơ bản
* Bao gồm tổng số công ty thỏa điều kiện filter

**GET /api/companies/{id}**

Endpoint lấy chi tiết một công ty:

* Verify company ID tồn tại trong database
* Return đầy đủ thông tin của công ty bao gồm danh sách jobs đã đăng
* Nếu không tìm thấy, trả về 404 Not Found

**PUT /api/companies/{id}**

Endpoint cập nhật thông tin công ty:

* Verify người dùng có quyền update (company\_admin của công ty đó)
* Validate input data
* Update thông tin trong MongoDB
* Return thông tin công ty sau khi cập nhật

**DELETE /api/companies/{id}**

Endpoint xóa công ty:

* Chỉ admin hệ thống mới có quyền xóa
* Xóa tất cả jobs liên quan đến công ty
* Xóa company document từ MongoDB
* Return success response

**POST /api/companies/{id}/jobs**

Endpoint đăng job mới cho công ty:

* Verify người dùng là company\_admin của công ty đó
* Validate thông tin job (title, mô tả, requirements, etc)
* Tạo Job object và liên kết với company
* Insert vào MongoDB collection jobs
* Return thông tin job đã tạo

## Kết luận

Chương 3 đã trình bày chi tiết quá trình triển khai hệ thống quản lý tuyển dụng từ kiến trúc tổng thể đến các API endpoints cụ thể. Hệ thống được xây dựng trên nền tảng kiến trúc 3 lớp với ASP.NET Core Web API, đảm bảo sự tách biệt rõ ràng giữa presentation, business logic và data access layers. Việc áp dụng Clean Architecture kết hợp với Repository Pattern không chỉ giúp code dễ maintain và test mà còn cho phép linh hoạt thay đổi implementation mà không ảnh hưởng đến toàn bộ hệ thống.

Các công nghệ được lựa chọn đều phục vụ mục đích cụ thể: MongoDB đáp ứng yêu cầu lưu trữ dữ liệu linh hoạt với schema động, Redis tối ưu hiệu năng thông qua caching strategy, JWT đảm bảo stateless authentication với refresh token rotation tăng cường bảo mật. Đặc biệt, Cache Aside Pattern và query optimization strategies được triển khai nhằm giảm thiểu load lên database và cải thiện response time. Các best practices về security như input validation, HTTPS enforcement, CORS configuration được tuân thủ nghiêm ngặt để bảo vệ hệ thống khỏi các mối đe dọa bảo mật phổ biến.

Qua quá trình cài đặt và cấu hình môi trường, các API endpoints được triển khai đầy đủ cho các nghiệp vụ chính: authentication, user management, job management và application management. Mỗi endpoint được thiết kế với validation chặt chẽ, error handling đầy đủ và logging chi tiết để dễ dàng troubleshooting. Với kiến trúc được thiết kế kỹ lượng và các công nghệ phù hợp, hệ thống đã sẵn sàng để triển khai và vận hành ổn định trong môi trường production.

# Kết luận

TÀI LIỆU THAM KHẢO

1. Redis Commands Documentation: <https://redis.io/docs/latest/commands/>.
2. Redis Official Website: <https://redis.io/>.
3. Contabo Blog: Everything about Redis:<https://contabo.com/blog/everything-about-redis/>.
4. Codersee: Redis Database Explained: <https://codersee.com/redis-database-explained/>.